


APPLICATION FOR UNITED STATES LETTERS PATENT

FOR

**EXPRESSION OF FRUCTOSE 1,6 BISPHTHOSPHATE
ALDOLASE IN TRANSGENIC PLANTS**

BY

GERARD BARRY
NORDINE CHEIKH
GANESH KISHORE

<p>EXPRESS MAIL MAILING LABEL</p> <p>NUMBER EI371156168US</p> <p>DATE OF DEPOSIT June 16, 1998</p> <p>I hereby certify that this paper or fee is being deposited with the United States Postal Service "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington D.C. 20231.</p> <p> _____ Signature</p>
--

This application is based on U.S. Provisional Application Serial No. 60/049,995, filed June 17, 1997.

FIELD OF THE INVENTION

This invention relates to the expression of fructose 1,6 bisphosphate aldolase (FDA) in transgenic plants to increase or improve plant growth and development, yield, vigor, stress tolerance, carbon allocation and storage into various storage pools, and distribution of starch. Transgenic plants expressing FDA have increased carbon assimilation, export and storage in plant source and sink organs, which results in growth, yield and quality improvements in crop plants.

BACKGROUND OF THE INVENTION

Recent advances in genetic engineering have provided the prerequisite tools to transform plants to contain alien (often referred to as "heterologous") or improved endogenous genes. These genes can lead either to an improvement of an already existing pathway in plant tissues or to an introduction of a novel pathway to modify product levels, increase metabolic efficiency, and or save on energy cost to the cell. It is presently possible to produce plants with unique physiological and biochemical traits and characteristics of high agronomic and crop processing importance. Traits that play an essential role in plant growth and development, crop yield potential and stability, and crop quality and composition include enhanced carbon assimilation, efficient carbon storage, and increased carbon export and partitioning.

Atmospheric carbon fixation (photosynthesis) by plants represents the major source of energy to support processes in all living organisms. The primary sites of photosynthetic activity, generally referred to as "source organs", are mature leaves and, to a lesser extent, green stems. The major carbon products of source leaves are starch, which represents the transitory storage form of carbohydrate in the chloroplast, and sucrose, which represents the predominant form of carbon transport in higher plants. Other plant parts named "sink organs" (e.g., roots, fruit, flowers, seeds, tubers, and bulbs) are generally not autotrophic and depend on import of sucrose or other major translocatable carbohydrates for their growth and development. The storage sinks

deposit the imported metabolites as sucrose and other oligosaccharides, starch and other polysaccharides, proteins, and triglycerides.

In leaves, the primary products of the Calvin Cycle (the biochemical pathway leading to carbon assimilation) are glyceraldehyde 3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP), also known as triose phosphates (triose-P). The condensation of G3P and DHAP into fructose 1,6 biphosphate (FBP) is catalyzed reversibly by the enzyme fructose 1,6 biphosphate aldolase (FDA), and various isozymes are known. The acidic isoenzyme appears to be chloroplastic and comprises about 85% of the total leaf aldolase activity. The basic isoenzyme is cytosolic. Both isoenzymes appear to be encoded by the nuclear genome and are encoded by different genes (Lebherz et al., 1984).

In the leaf, the chloroplast FDA is an essential enzyme in the Calvin Cycle, where its activity generates metabolites for starch biosynthesis. Removal of more than 40% of the plastidic aldolase enzymatic activity by antisense technology reduced leaf starch accumulation as well as soluble proteins and chlorophyll levels but also reduced plant growth and root formation (Sonnewald et al., 1994). In contrast, the cytosolic FDA is part of the sucrose biosynthetic pathway where it catalyzes the reaction of FBP production. Moreover, cytosolic FDA is also a key enzyme in the glycolytic and gluconeogenesis pathways in both source and sink plant tissues.

In the potato industry, production of higher starch and uniform solids tubers is highly desirable and valuable. The current potato varieties that are used for french fry production, such as Russet Burbank and Shepody, suffer from a non-uniform deposition of solids between the tuber pith (inner core) and the cortex (outer core). French fry strips that are taken from pith tissue are higher in water content when compared to outer cortex french fry strips; cortex tissue typically displays a solids level of twenty-four percent whereas pith tissue typically displays a solids level of seventeen percent. Consequently, in the french fry production process, the pith strips need to be blanched, dried, and par-fried for longer times to eliminate the excess water. Adequate processing of the pith fries results in the over-cooking of fries from the high solids cortex. The blanching, drying, and par frying times of the french fry processor need to be adjusted accordingly to accommodate the low solids pith strips and the high solids cortex strips. A higher solids potato with a more uniform distribution of starch from pith to cortex would allow

for a more uniform finished fry product, with higher plant throughput and cost savings due to reduced blanch, dry and par-fry times.

Although various fructose 1,6 biphosphate aldolases have been previously characterized, it has been discovered that overexpression of the enzyme in a transgenic plant provides
5 advantageous results in the plant such as increasing the assimilation, export and storage of carbon, increasing the production of oils and/or proteins in the plant and improving tuber solids uniformity.

SUMMARY OF THE INVENTION

10 The present invention provides structural DNA constructs that encode a fructose 1,6 biphosphate aldolase (FDA) enzyme and that are useful in increasing carbon assimilation, export, and storage in plants.

In accomplishing the foregoing, there is provided, in accordance with one aspect of the present invention, a method of producing genetically transformed plants that have elevated
15 carbon assimilation, storage, export, and improved solids uniformity comprising the steps of:

- (a) Inserting into the genome of a plant a recombinant, double-stranded DNA molecule comprising
 - (i) a promoter that functions in the cells of a target plant tissue,
 - (ii) a structural DNA sequence that causes the production of an RNA sequence that
20 encodes a fructose 1,6 biphosphate aldolase enzyme,
 - (iii) a 3' non-translated DNA sequence that functions in plant cells to cause transcriptional termination and the addition of polyadenylated nucleotides to the 3' end of the RNA sequence;
- (b) obtaining transformed plant cells; and
- 25 (c) regenerating from transformed plant cells genetically transformed plants that have elevated FDA activity.

In another aspect of the present invention there is provided a recombinant, double-stranded DNA molecule comprising in sequence

- (i) a promoter that functions in the cells of a target plant tissue,

- (ii) a structural DNA sequence that causes the production of an RNA sequence that encodes a fructose 1,6 biphosphate aldolase enzyme,
(iii) a 3' non-translated DNA sequence that functions in plant cells to cause transcriptional termination and the addition of polyadenylated nucleotides to the 3' end of the RNA sequence.

In a further aspect of the present invention, the structural DNA sequence that causes the production of an RNA sequence that encodes a fructose 1,6 biphosphate aldolase enzyme is coupled with a chloroplast transit peptide to facilitate transport of the enzyme to the plastid.

In accordance with the present invention, an improved means for increasing carbon assimilation, storage and export in the source tissues of various plants is provided. Further means of improved carbon accumulation in sinks (such as roots, tubers, seeds, stems, and bulbs) are provided, thus increasing the size of various sinks (larger roots, tubers, etc.) and subsequently increasing yield and crop productivity. The increased carbon availability to these sinks would also improve composition and use efficiency in the sink (oil, protein, starch and/or sucrose production, and/or solids uniformity).

Various advantages may be achieved by the aims of the present invention, including:

First, increasing the expression of the FDA enzyme in the chloroplast would increase the flow of carbon through the Calvin Cycle and increase atmospheric carbon assimilation during early photoperiod. This would result in an increase in photosynthetic efficiency and an increase in chloroplast starch production (a leaf carbon storage form degraded during periods when photosynthesis is low or absent). Both of these responses would lead to an increase in sucrose production by the leaf and a net increase in carbon export during a given photoperiod. This increase in source capacity is a desirable trait in crop plants and would lead to increased plant growth, storage ability, yield, vigor, and stress tolerance.

Second, increasing FDA expression in the cytosol of photosynthetic cells would lead to an increase in sucrose production and export out of source leaves. This increase in source capacity is a desirable trait in crop plants and would lead to increased plant growth, storage ability, yield, vigor, and stress tolerance.

Third, expression of FDA in sink tissues can show several desirable traits, such as increased amino acid and/or fatty acid pools via increases in carbon flux through glycolysis (and

thus pyruvate levels) in seeds or other sinks and increased starch levels as result of increased production of glucose 6-phosphate in seeds, roots, stems, and tubers where starch is a major storage nonstructural carbohydrate (reverse glycolysis). This increase in sink strength is a desirable trait in crop plants and would lead to increased plant growth, storage ability, yield, vigor, and stress tolerance.

Fourth, the invention is particularly desirable for use in the commercial production of foods derived from potatoes. Potatoes used for the production of french fries and other products suffer from a non-uniform distribution of solids between the tuber pith (inner core) and the cortex (outer core). Thus, french fry strips from the pith regions of such tubers have a low solids content and a high water content in comparison to cortex strips from the same tubers. Therefore, the french fry processor attempts to adjust the processing parameters so that the final inner strips are sufficiently cooked while the outer cortex strips are not overcooked. The results of such adjustments, however, are highly variable and may lead to poor quality product. Transgenic potatoes expressing *fda* will provide to the french-fry and potato chip processor a raw product that consistently displays a higher tuber solids uniformity with acceptable agronomic traits. In the french fry plant production process, inner pith fry strips from higher solids uniformity tubers will require less time to blanch, less time to dry to a specific solids content, and less time to par-fry before freezing and shipping to retail and institutional end-users.

Therefore, with respect to potatoes, the present invention provides 1) a higher quality, more uniform finish fry product in which french fries from all tuber regions, when processed, are nearly the same, 2) a higher through-put in the french fry processing plant due to lower processing times, and 3) processor cost savings due to lower energy input required for lower blanch, dry, and par-fry times. A raw tuber product that displays a higher solids uniformity will also produce a potato chip that has a reduced saddle curl, and a reduced tendency for center bubble, which are undesirable qualities in the potato chip industry. Reduced fat content would also result; this would contribute to improved consumer appeal and lower oil use (and costs) for the processor. The increase in solids uniformity will also translate to an increase in overall tuber solids. For both the french fry and chipping industries, this overall tuber solids increase will also result in higher through-put in the processing plant due to lower processing times, and cost savings due to lower energy input for blanching, drying, par-frying, and finish frying.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the nucleotide sequence and deduced amino acid sequence of a fructose 1,6 biphosphate aldolase gene from *E. coli* (SEQ ID No:1).

5 **Figure 2** shows a plasmid map for plant transformation vector pMON17524.

Figure 3 shows a plasmid map for plant transformation vector pMON17542.

10 **Figure 4** shows the change in diurnal fluctuations of sucrose, glucose, and starch levels in tobacco leaves expressing the *fda* transgene (pMON17524) and control (pMON17227). The light period is from 7:00 to 19:00 hours. Only fully expanded and non-senescing leaves were sampled.

Figure 5 shows a plasmid map for plant transformation vector pMON13925.

15

Figure 6 shows a plasmid map for plant transformation vector pMON17590.

Figure 7 shows a plasmid map for plant transformation vector pMON13936.

20 **Figure 8** shows a plasmid map for plant transformation vector pMON17581.

Figure 9 shows potato tuber cross-sections of improved solids uniformity Segal Russet Burbank lines (top row) versus unimproved nontransgenic Russet Burbank (bottom row).

25

DETAILED DESCRIPTION OF THE INVENTION

This invention is directed to a method for producing plant cells and plants demonstrating an increased or improved growth and development, yield, quality, starch storage uniformity, vigor, and/or stress tolerance. The method utilizes a DNA sequence encoding an *fda* (fructose 1,6 biphosphate aldolase) gene integrated in the cellular genome of a plant as the result of

genetic engineering and causes expression of the FDA enzyme in the transgenic plant so produced. Plants that overexpress the FDA enzyme exhibit increased carbon flow through the Calvin Cycle and increased atmospheric carbon assimilation during early photoperiod resulting in an increase in photosynthetic efficiency and an increase in starch production. Thus, such plants exhibit higher levels of sucrose production by the leaf and the ability to achieve a net increase in carbon export during a given photoperiod. This increase in source capacity leads to increased plant growth that in turn generates greater biomass and/or increases the size of the sink and ultimately providing greater yields of the transgenic plant. This greater biomass or increased sink size may be evidenced in different ways or plant parts depending on the particular plant species or growing conditions of the plant overexpressing the FDA enzyme. Thus, increased size resulting from overexpression of FDA may be seen in the seed, fruit, stem, leaf, tuber, bulb or other plant part depending upon the plant species and its dominant sink during a particular growth phase and upon the environmental effects caused by certain growing conditions, e.g. drought, temperature or other stresses. Transgenic plants overexpressing FDA may therefore have increased carbon assimilation, export and storage in plant source and sink organs, which results in growth, yield, and uniformity and quality improvements.

Plants overexpressing FDA may also exhibit desirable quality traits such as increased production of starch, oils and/or proteins depending upon the plant species overexpressing the FDA. Thus, overexpression of FDA in a particular plant species may affect or alter the direction of the carbon flux thereby directing metabolite utilization and storage either to starch production, protein production or oil production via the role of FDA in the glycolysis and gluconeogenesis metabolic pathways.

The mechanism whereby the expression of exogenous FDA modifies carbon relationships is believed to derive from source-sink relationships. The leaf tissue is a sucrose source, and if more sucrose resulting from the activity of increased FDA expression is transported to a sink, it results in increased storage carbon (sugars, starch, oil, protein, etc.) or nitrogen (protein, etc.) per given weight of the sink tissue.

The expression in a plant of a gene that exists in double-stranded DNA form involves transcription of messenger RNA (mRNA) from one strand of the DNA by RNA polymerase enzyme, and the subsequent processing of the mRNA primary transcript inside the nucleus. This processing involves a 3' non-translated region, which adds polyadenylate nucleotides to the 3' end of the RNA. Transcription of DNA into mRNA is regulated by a region of DNA usually referred to as the promoter. The promoter region contains a sequence of bases that signals RNA polymerase to associate with the DNA and to initiate the transcription of mRNA using one of the DNA strands as a template to make a corresponding complimentary strand of RNA. This RNA is then used as a template for the production of the protein encoded therein by the cells protein biosynthetic machinery.

A number of promoters that are active in plant cells have been described in the literature. These include the nopaline synthase (NOS) and octopine synthase (OCS) promoters (which are carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*), the caulimovirus promoters such as the cauliflower mosaic virus (CaMV) 19S and 35S and the figwort mosaic virus (FMV) 35S-promoters, the light-inducible promoter from the small subunit of ribulose-1,5-bisphosphate carboxylase (ssRUBISCO), a very abundant plant polypeptide, and the chlorophyll a/b binding protein gene promoters, etc. All of these promoters have been used to create various types of DNA constructs that have been expressed in plants; see, e.g., PCT publication WO 84/02913.

Promoters that are known to or are found to cause transcription of DNA in plant cells can be used in the present invention. Such promoters may be obtained from a variety of sources such as plants and plant viruses and include, but are not limited to, the enhanced CaMV35S promoter and promoters isolated from plant genes such as ssRUBISCO genes. As described below, it is preferred that the particular promoter selected should be capable of causing sufficient expression to result in the production of an effective amount of fructose 1,6 bisphosphate aldolase enzyme to cause the desired increase in carbon assimilation, export or storage. Expression of the double-stranded DNA molecules of the present invention can be driven by a constitutive promoter, expressing the DNA molecule in all or most of the tissues of the plant. Alternatively, it may be preferred to cause expression of the *fda* gene in specific tissues of the plant, such as leaf, stem, root, tuber, seed, fruit, etc. The promoter chosen will have the desired tissue and developmental specificity. Those skilled in the art will recognize that the amount of fructose 1,6 bisphosphate

aldolase needed to induce the desired increase in carbon assimilation, export, or storage may vary with the type of plant. Therefore, promoter function should be optimized by selecting a promoter with the desired tissue expression capabilities and approximate promoter strength and selecting a transformant that produces the desired fructose 1,6 biphosphate aldolase activity or the desired change in metabolism of carbohydrates in the target tissues. This selection approach from the pool of transformants is routinely employed in expression of heterologous structural genes in plants because there is variation between transformants containing the same heterologous gene due to the site of gene insertion within the plant genome (commonly referred to as "position effect"). In addition to promoters that are known to cause transcription (constitutively or tissue-specific) of DNA in plant cells, other promoters may be identified for use in the current invention by screening a plant cDNA library for genes that are selectively or preferably expressed in the target tissues of interest and then isolating the promoter regions by methods known in the art. In particular, it may be desirable to use a bundle sheath cell specific (or cell enhanced expression) promoter for use with C4 plants such as corn, sorghum, and sugarcane to obtain the yield benefits of overexpression of FDA and not use a constitutive promoter or a promoter with mesophyll cell enhanced expression properties.

For the purpose of expressing the *fda* gene in source tissues of the plant, such as the leaf or stem, it is preferred that the promoters utilized in the double-stranded DNA molecules of the present invention have relatively high expression in these specific tissues. For this purpose, one may also choose from a number of promoters for genes with leaf-specific or leaf-enhanced expression. Examples of such genes known from the literature are the chloroplast glutamine synthetase GS2 from pea (Edwards et al., 1990), the chloroplast fructose-1,6-bisphosphatase (FBPase) from wheat (Lloyd et al., 1991), the nuclear photosynthetic ST-LS1 from potato (Stockhaus et al., 1989), and the phenylalanine ammonia-lyase (PAL) and chalcone synthase (CHS) genes from *Arabidopsis thaliana* (Leyva et al., 1995). Also shown to be active in photosynthetically active tissues are the ribulose-1,5-bisphosphate carboxylase (RUBISCO), isolated from eastern larch (*Larix laricina*) (Campbell et al., 1994); the *cab* gene, encoding the chlorophyll a/b-binding protein of PSII, isolated from pine (*cab6*; Yamamoto et al., 1994), wheat (*Cab-1*; Fejes et al., 1990), spinach (*CAB-1*; Luebberstedt et al., 1994), and rice (*cab1R*; Luan et al., 1992); the pyruvate orthophosphate dikinase (PPDK) from maize (Matsuoka et al., 1993); the

tobacco Lhcb1*2 gene (Cerdan et al., 1997); the *Arabidopsis thaliana* SUC2 sucrose-H⁺ symporter gene (Truernit et al., 1995); and the thylacoid membrane proteins, isolated from spinach (psaD, psaF, psaE, PC, FNR, atpC, atpD, cab, rbcS; Oelmueller et al., 1992). Other chlorophyll a/b-binding proteins have been studied and described in the literature, such as Lhcb and PsbP from white mustard (*Sinapis alba*; Kretsch et al., 1995). Homologous promoters to those described here may also be isolated from and tested in the target or related crop plant by standard molecular biology procedures.

For the purpose of expressing the *fda* in sink tissues of the plant, for example the tuber of the potato plant; the fruit of tomato; or seed of maize, wheat, rice, or barley, it is preferred that the promoters utilized in the double-stranded DNA molecules of the present invention have relatively high expression in these specific tissues. A number of genes with tuber-specific or tuber-enhanced expression are known, including the class I patatin promoter (Bevan et al., 1986; Jefferson et al., 1990); the potato tuber ADPGPP genes, both the large and small subunits (Muller et al., 1990); sucrose synthase (Salanoubat and Belliard, 1987, 1989); the major tuber proteins including the 22 kDa protein complexes and proteinase inhibitors (Hannapel, 1990); the granule bound starch synthase gene (GBSS) (Rohde et al., 1990); and the other class I and II patatins (Rocha-Sosa et al., 1989; Mignery et al., 1988). Other promoters can also be used to express a fructose 1,6 biphosphate aldolase gene in specific tissues, such as seeds or fruits. The promoter for β -conglycinin (Tierney, 1987) or other seed-specific promoters, such as the napin and phaseolin promoters, can be used to over-express an *fda* gene specifically in seeds. The zeins are a group of storage proteins found in maize endosperm. Genomic clones for zein genes have been isolated (Pedersen et al., 1982), and the promoters from these clones, including the 15 kDa, 16 kDa, 19 kDa, 22 kDa, 27 kDa, and gamma genes, could also be used to express an *fda* gene in the seeds of maize and other plants. Other promoters known to function in maize, wheat, or rice include the promoters for the following genes: *waxy*, *Brittle*, *Shrunken 2*, branching enzymes I and II, starch synthases, debranching enzymes, oleosins, glutelins, and sucrose synthases. Particularly preferred promoters for maize endosperm expression, as well as in wheat and rice, of an *fda* gene is the promoter for a glutelin gene from rice, more particularly the Osgt-1 promoter (Zheng et al., 1993); the maize granule-bound starch synthase (*waxy*) gene (zmGBS); the rice small subunit ADPGPP promoter (osAGP); and the zein promoters, particularly the

maize 27 kDa zein gene promoter (zm27) (see, generally, Russell et al., 1997). Examples of promoters suitable for expression of an *fda* gene in wheat include those for the genes for the ADPglucose pyrophosphorylase (ADPGPP) subunits, for the granule bound and other starch synthases, for the branching and debranching enzymes, for the embryogenesis-abundant proteins, for the gliadins, and for the glutenins. Examples of such promoters in rice include those for the genes for the ADPGPP subunits, for the granule bound and other starch synthases, for the branching enzymes, for the debranching enzymes, for sucrose synthases, and for the glutelins. A particularly preferred promoter is the promoter for rice glutelin, Osgt-1. Examples of such promoters for barley include those for the genes for the ADPGPP subunits, for the granule bound and other starch synthases, for the branching enzymes, for the debranching enzymes, for sucrose synthases, for the hordeins, for the embryo globulins, and for the aleurone-specific proteins.

The solids content of root tissue may be increased by expressing an *fda* gene behind a root-specific promoter. An example of such a promoter is the promoter from the acid chitinase gene (Samac et al., 1990). Expression in root tissue could also be accomplished by utilizing the root-specific subdomains of the CaMV35S promoter that have been identified (Benfey et al., 1989).

The RNA produced by a DNA construct of the present invention may also contain a 5' non-translated leader sequence. This sequence can be derived from the promoter selected to express the gene and can be specifically modified so as to increase translation of the mRNA. The 5' non-translated regions can also be obtained from viral RNAs, from suitable eukaryotic genes, or from a synthetic gene sequence. The present invention is not limited to constructs, as presented in the following examples, wherein the non-translated region is derived from the 5' non-translated sequence that accompanies the promoter sequence. Rather, the non-translated leader sequence can be derived from an unrelated promoter or coding sequence.

In monocots, an intron is preferably included in the gene construct to facilitate or enhance expression of the coding sequence. Examples of suitable introns include the HSP70 intron and the rice actin intron, both of which are known in the art. Another suitable intron is the castor bean catalase intron (Suzuki et al., 1994)

Polyadenylation signal

The 3' non-translated region of the chimeric plant gene contains a polyadenylation signal that functions in plants to cause the addition of polyadenylate nucleotides to the 3' end of the RNA. Examples of suitable 3' regions are (1) the 3' transcribed, non-translated regions
5 containing the polyadenylation signal of *Agrobacterium* tumor-inducing (Ti) plasmid genes, such as the nopaline synthase (NOS) gene, and (2) plant genes like the soybean storage protein genes and the small subunit of the ribulose-1,5-bisphosphate carboxylase (ssRUBISCO) gene.

Plastid-directed expression of fructose-1,6-bisphosphate aldolase activity

10 In one embodiment of the invention, the *fda* gene may be fused to a chloroplast transit peptide, in order to target the FDA protein to the plastid. As used hereinafter, chloroplast and plastid are intended to include the various forms of plastids including amyloplasts. Many plastid-localized proteins are expressed from nuclear genes as precursors and are targeted to the plastid by a chloroplast transit peptide (CTP), which is removed during the import steps.
15 Examples of such chloroplast proteins include the small subunit of ribulose-1,5-bisphosphate carboxylase (ssRUBISCO, SSU), 5-enolpyruvate shikimate-3-phosphate synthase (EPSPS), ferredoxin, ferredoxin oxidoreductase, the light-harvesting-complex protein I and protein II, and thioredoxin F. It has been demonstrated that non-plastid proteins may be targeted to the chloroplast by use of protein fusions with a CTP and that a CTP sequence is sufficient to target a
20 protein to the plastid. Those skilled in the art will also recognize that various other chimeric constructs can be made that utilize the functionality of a particular plastid transit peptide to import the fructose-1,6-diphosphate aldolase enzyme into the plant cell plastid. The *fda* gene could also be targeted to the plastid by transformation of the gene into the chloroplast genome (Daniell et al., 1998).

Fructose 1,6 bisphosphate aldolases

As used herein, the term "fructose 1, 6-bisphosphate aldolase" means an enzyme (E.C. 4.1.2.13) that catalyzes the reversible cleavage of fructose 1,6-bisphosphate to form glyceraldehyde 3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP). Aldolase enzymes
30 are divided into two classes, designated class I and class II (Witke and Gotz, 1993). Various *fda*

genes encoding the enzyme have been sequenced, as have numerous proteins, such as the cytosolic enzyme from maize (GenBank Accession S07789;S10638), cytosolic enzyme from rice (GenBank Accession JQ0543), cytosolic enzyme from spinach (GenBank Accession S31091;S22093), from *Arabidopsis thaliana* (GenBank Accession S11958), from spinach chloroplast (GenBank Accession S31090;A21815;S22092), from yeast (*S. cerevisiae*) (GenBank Accession S07855; S37882; S12945; S39178; S44523;X75781), from *Rhodobacter sphaeroides* (GenBank Accession B40767;D41080), from *B. subtilis* (GenBank Accession S55426; D32354; E32354; D41835), from garden pea (GenBank Accession S29048; S34411), from garden pea chloroplast (GenBank Accession S29047; S34410), from maize (GenBank Accession S05019), from *Chlamydomonas reinhardtii* (GenBank Accession S48639; S58485; S58486; S34367), from *Corynebacterium glutamicum* (GenBank Accession S09283; X17313), from *Campylobacter jejuni* (GenBank Accession S52413), from *Haemophilus influenzae* (strain Rd KW20) (GenBank Accession C64074), from *Streptococcus pneumonia* (GenBank Accession AJ005697), from rice (GenBank Accession X53130), and from the maize anaerobically regulated gene (GenBank Accession X12872).

The class I enzymes may be isolated from higher eukaryotes, such as animals and plants, and in some prokaryotes, including *Peptococcus aerogens*, (Lebherz and Rutter, 1973), *Lactobacillus casei* (London and Kline, 1973), *Escherichia coli* (Stribling and Perham, 1973), *Mycobacterium smegmatis* (Bai et al., 1975), and most staphylococcal species (Gotz et al., 1979). The gene for the FDA enzyme may be obtained by known methods and has already been done so for several organisms, such as rabbit (Lai et al., 1974), human (Besmond et al., 1983), rat (Tsutsumi et al., 1984), *Trypanosoma brucei* (Clayton, 1985), and *Arabidopsis thaliana* (Chopra et al., 1990). These class I enzymes are invariably tetrameric proteins with a total molecular weight of about 160 kDa and function by imine formation between the substrate and a lysine residue in the active site (Alfounder et al., 1989).

In animal, three class I isozymes, classified as A, B, and C, are expressed in the cytosol of muscle, liver, and brain tissue respectively, and they differ from plant aldolases in their expression and compartmentation patterns (Joh et al., 1986). In the leaves of higher plants, FDA is a class I enzyme, and two different isoenzymes within the class have been documented. One is contained in the chloroplast and the other in the cytosol (Lebherz et al., 1984). The acidic plant

isozyme appear to be chloroplastic and comprises about 85% of the total leaf aldolase activity. The basic plant isozyme is cytosolic, and both isozymes appear to be encoded by the nuclear genome and are encoded by different genes (Lebherz et al., 1984).

The class II type aldolases are normally dimeric with molecular mass of approximately 80 kDa, and their activity depends on divalent metal ions. The class II enzymes may be isolated from prokaryotes, such as blue-green algae and bacteria, and eukaryotic green algae and fungi (Baldwin et al., 1978). The gene for the FDA class II enzyme may be obtained by known methods and has already been done so from several organisms including *Saccharomyces cerevisiae* (Jack and Harris, 1971), *Bacillus stearothermophilus* (Jack, 1973), and *Escherichia coli* (Baldwin et al., 1978).

It is believed that highly homologous class II fructose 1, 6-bisphosphate aldolases with similar catalyzing activity will also be found in other species of microorganism, such as *Saccharomyces* (*Saccharomyces cerevisiae*); *Bacillus* (*Bacillus subtilis*); *Rhodobacter* (*Rhodobacter sphaeroides*); *Plasmodium* (*Plasmodium falciparum*, *Plasmodium berghei*); *Trypanosoma* (*Trypanosoma brucei*); *Chlamydomonas* (*Chlamydomonas reinhardtii*); *Candida* (*Candida albicans*); *Corynebacterium* (*Corynebacterium glutamicum*); *Campylobacter* (*Campylobacter jejuni*); and *Haemophilus* (*Haemophilus influenza*).

Such sequences can be readily isolated by methods well known in the art, for example by nucleic acid hybridization. The hybridization properties of a given pair of nucleic acids are an indication of their similarity or identity. Nucleic acid sequences can be selected on the basis of their ability to hybridize with known *fda* sequences. Low stringency conditions may be used to select sequences with less homology or identity. One may wish to employ conditions such as about 0.15 M to about 0.9 M sodium chloride, at temperatures ranging from about 20°C to about 55°C. High stringency conditions may be used to select for nucleic acid sequences with higher degrees of identity to the disclosed sequences. Conditions typically employed may include about 0.02 M to about 0.15 M sodium chloride, about 0.5% to about 5% casein, about 0.02% SDS or about 0.1% N-laurylsarcosine, about 0.001 M to about 0.03 M sodium citrate, at hybridization temperatures between about 50°C and about 70°C. More preferably, high stringency conditions are about 0.02 M sodium chloride, about 0.5% casein, about 0.02% SDS, about 0.001 M sodium

citrate, at a temperature of about 50°C. The skilled individual will recognize that numerous variations are possible in the conditions and means by which nucleic acid hybridization can be performed to isolate *fda* sequences having similarity to *fda* sequences known in the art and are not limited to those explicitly disclosed herein. Preferably, such an approach is used to isolate *fda* sequences having greater than about 60% identity with the disclosed *E.coli fda* sequence, more preferably greater than about 70% identity, most preferably greater than about 80% identity.

Depending on growth conditions *Euglena gracilis*, *Chlamydomonas mundana*, and *Chlamydomonas reinhardtii* produce either a class I or a class II aldolase (Cremona, 1968; Russell and Gibbs, 1967; Guerrini et al., 1971).

The isolation of a class II *fda* gene from *E. coli* is described in the following examples. Its DNA sequence is given as SEQ ID NO:1 and shown in Figure 1. The amino acid sequence is shown in SEQ ID NO:2 and shown in Figure 1. This gene can be used as isolated by inserting it into plant expression vectors suitable for the transformation method of choice as described. The *E. coli* FDA enzyme has an apparent pH optimum range near pH 7-9 and retains activity in the lower pH range of 5-7 (Baldwin et al., 1978; Alfounder et al., 1989).

Thus, many different genes that encode a fructose 1,6 bisphosphate aldolase activity may be isolated and used in the present invention.

Synthetic gene construction

A carbohydrate metabolizing enzyme considered in this invention includes any sequence of amino acids, such as protein, polypeptide, or peptide fragment, that demonstrates the ability to catalyze a reaction involved in the synthesis or degradation of starch or sucrose. These can be sequences obtained from a heterologous source, such as algae, bacteria, fungi, and protozoa, or endogenous plant sequences, by which is meant any sequence that can be naturally found in a plant cell, including native (indigenous) plant sequences as well as sequences from plant viruses or plant pathogenic bacteria.

It will be recognized by one of ordinary skill in the art that carbohydrate metabolizing enzyme gene sequences may also be modified using standard techniques such as site-specific mutation or PCR, or modification of the sequence may be accomplished by producing a synthetic

nucleic acid sequence and will still be considered a carbohydrate biosynthesis enzyme nucleic acid sequence of this invention. For example, “wobble” positions in codons may be changed such that the nucleic acid sequence encodes the same amino acid sequence, or alternatively, codons can be altered such that conservative amino acid substitutions result. In either case, the peptide or protein maintains the desired enzymatic activity and is thus considered part of this invention.

A nucleic acid sequence to a carbohydrate metabolizing enzyme may be a DNA or RNA sequence, derived from genomic DNA, cDNA, mRNA, or may be synthesized in whole or in part. The structural gene sequences may be cloned, for example, by isolating genomic DNA from an appropriate source and amplifying and cloning the sequence of interest using a polymerase chain reaction (PCR). Alternatively, the gene sequences may be synthesized, either completely or in part, especially where it is desirable to provide plant-preferred sequences. Thus, all or a portion of the desired structural gene may be synthesized using codons preferred by a selected plant host. Plant-preferred codons may be determined, for example, from the codons used most frequently in the proteins expressed in a particular plant host species. Other modifications of the gene sequences may result in mutants having slightly altered activity.

If desired, the gene sequence of the *fda* gene can be changed without changing the protein sequence in such a manner as may increase expression and thus even more positively affect carbohydrate content in transformed plants. A preferred manner for making the changes in the gene sequence is set out in PCT Publication WO 90/10076. A gene synthesized by following the methodology set out therein may be introduced into plants as described below and result in higher levels of expression of the FDA enzyme. This may be particularly useful in monocots such as maize, rice, wheat, sugarcane, and barley.

Combinations with other transgenes

The effect of *fda* in transgenic plants may be enhanced by combining it with other genes that positively affect carbohydrate assimilation or content, such as a gene encoding for a sucrose phosphorylase as described in PCT Publication WO 96/24679, or ADPGPP genes such as the *E. coli glgC* gene and its mutant *glgC16*. PCT Publication WO 91/19806 discloses how to incorporate the latter gene into many plant species in order to increase starch or solids. Another gene that can be combined with *fda* to increase carbon assimilation, export or storage is a gene

encoding for sucrose phosphate synthase (SPS). PCT Publication WO 92/16631 discloses one such gene and its use in transgenic plants.

Plant transformation/regeneration

5 In developing the nucleic acid constructs of this invention, the various components of the construct or fragments thereof will normally be inserted into a convenient cloning vector, e.g., a plasmid that is capable of replication in a bacterial host, e.g., *E. coli*. Numerous vectors exist that have been described in the literature, many of which are commercially available. After each cloning, the cloning vector with the desired insert may be isolated and subjected to further
10 manipulation, such as restriction digestion, insertion of new fragments or nucleotides, ligation, deletion, mutation, resection, etc. so as to tailor the components of the desired sequence. Once the construct has been completed, it may then be transferred to an appropriate vector for further manipulation in accordance with the manner of transformation of the host cell.

A recombinant DNA molecule of the invention typically includes a selectable marker so
15 that transformed cells can be easily identified and selected from non-transformed cells. Examples of such include, but are not limited to, a neomycin phosphotransferase (nptII) gene (Potrykus et al., 1985), which confers kanamycin resistance. Cells expressing the nptII gene can be selected using an appropriate antibiotic such as kanamycin or G418. Other commonly used selectable markers include the *bar* gene, which confers bialaphos resistance; a mutant EPSP
20 synthase gene (Hinchee et al., 1988), which confers glyphosate resistance; a nitrilase gene, which confers resistance to bromoxynil (Stalker et al., 1988); a mutant acetolactate synthase gene (ALS), which confers imidazolinone or sulphonylurea resistance (European Patent Application 154,204, 1985); and a methotrexate resistant DHFR gene (Thillet et al., 1988).

Plants that can be made to have enhanced carbon assimilation, increased carbon export
25 and partitioning by practice of the present invention include, but are not limited to, *Acacia*, alfalfa, aneth, apple, apricot, artichoke, arugula, asparagus, avocado, banana, barley, beans, beet, blackberry, blueberry, broccoli, brussels sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, celery, cherry, cilantro, citrus, clementines, coffee, corn, cotton, cucumber, Douglas fir, eggplant, endive, escarole, eucalyptus, fennel, figs, gourd, grape, grapefruit, honey dew,

jicama, kiwifruit, lettuce, leeks, lemon, lime, Loblolly pine, mango, melon, mushroom, nut, oat, oil seed rape, okra, onion, orange, an ornamental plant, papaya, parsley, pea, peach, peanut, pear, pepper, persimmon, pine, pineapple, plantain, plum, pomegranate, poplar, potato, pumpkin, quince, radiata pine, radicchio, radish, raspberry, rice, rye, sorghum, Southern pine, soybean, spinach, squash, strawberry, sugarbeet, sugarcane, sunflower, sweet potato, sweetgum, tangerine, tea, tobacco, tomato, triticale, turf, a vine, watermelon, wheat, yams, and zucchini.

A double-stranded DNA molecule of the present invention containing an *fda* gene can be inserted into the genome of a plant by any suitable method. Suitable plant transformation vectors include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed, e.g., by Herrera-Estrella et al. (1983), Bevan (1984), Klee et al. (1985) and EPO publication 120,516. In addition to plant transformation vectors derived from the Ti or root-inducing (Ri) plasmids of *Agrobacterium*, alternative methods can be used to insert the DNA constructs of this invention into plant cells. Such methods may involve, for example, the use of liposomes, electroporation, chemicals that increase free DNA uptake, free DNA delivery via microprojectile bombardment, and transformation using viruses or pollen. DNA may also be inserted into the chloroplast genome (Daniell et al., 1998).

A plasmid expression vector suitable for the introduction of an *fda* gene in monocots using microprojectile bombardment is composed of the following: a promoter that is specific or enhanced for expression in the starch storage tissues in monocots, generally the endosperm, such as promoters for the zein genes found in the maize endosperm (Pedersen et al., 1982); an intron that provides a splice site to facilitate expression of the gene, such as the Hsp70 intron (PCT Publication WO93/19189); and a 3' polyadenylation sequence such as the nopaline synthase 3' sequence (NOS 3'; Fraley et al., 1983). This expression cassette may be assembled on high copy replicons suitable for the production of large quantities of DNA.

A particularly useful *Agrobacterium*-based plant transformation vector for use in transformation of dicotyledonous plants is plasmid vector pMON530 (Rogers et al., 1987). Plasmid pMON530 is a derivative of pMON505 prepared by transferring the 2.3 kb *Stu*I-*Hind*III fragment of pMON316 (Rogers et al., 1987) into pMON526. Plasmid pMON526 is a simple derivative of pMON505 in which the *Sma*I site is removed by digestion with *Xma*I, treatment with Klenow polymerase and ligation. Plasmid pMON530 retains all the properties of

pMON505 and the CaMV35S-NOS expression cassette and now contains a unique cleavage site for SmaI between the promoter and polyadenylation signal.

Binary vector pMON505 is a derivative of pMON200 (Rogers et al., 1987) in which the Ti plasmid homology region, LIH, has been replaced with a 3.8 kb HindIII to SmaI segment of the mini RK2 plasmid, pTJS75 (Schmidhauser and Helinski, 1985). This segment contains the RK2 origin of replication, oriV, and the origin of transfer, oriT, for conjugation into *Agrobacterium* using the tri-parental mating procedure (Horsch and Klee, 1986). Plasmid pMON505 retains all the important features of pMON200 including the synthetic multi-linker for insertion of desired DNA fragments, the chimeric NOS/NPTII/NOS gene for kanamycin resistance in plant cells, the spectinomycin/streptomycin resistance determinant for selection in *E. coli* and *A. tumefaciens*, an intact nopaline synthase gene for facile scoring of transformants and inheritance in progeny, and a pBR322 origin of replication for ease in making large amounts of the vector in *E. coli*. Plasmid pMON505 contains a single T-DNA border derived from the right end of the pTiT37 nopaline-type T-DNA. Southern blot analyses have shown that plasmid pMON505 and any DNA that it carries are integrated into the plant genome, that is, the entire plasmid is the T-DNA that is inserted into the plant genome. One end of the integrated DNA is located between the right border sequence and the nopaline synthase gene and the other end is between the border sequence and the pBR322 sequences.

Another particularly useful Ti plasmid cassette vector is pMON17227. This vector is described in PCT Publication WO 92/04449 and contains a gene encoding an enzyme conferring glyphosate resistance (denominated CP4), which is an excellent selection marker gene for many plants, including potato and tomato. The gene is fused to the *Arabidopsis* EPSPS chloroplast transit peptide (CTP2) and expressed from the FMV promoter as described therein.

When adequate numbers of cells (or protoplasts) containing the *fda* gene or cDNA are obtained, the cells (or protoplasts) are regenerated into whole plants. Choice of methodology for the regeneration step is not critical, with suitable protocols being available for hosts from Leguminosae (alfalfa, soybean, clover, etc.), Umbelliferae (carrot, celery, parsnip), Cruciferae (cabbage, radish, canola/rapeseed, etc.), Cucurbitaceae (melons and cucumber), Gramineae (wheat, barley, rice, maize, etc.), Solanaceae (potato, tobacco, tomato, peppers), various floral crops, such as sunflower, and nut-bearing trees, such as almonds, cashews, walnuts, and pecans.

See, e.g., Ammirato et al. (1984); Shimamoto et al. (1989); Fromm (1990); Vasil et al. (1990); Vasil et al. (1992); Hayashimoto (1990); and Datta et al. (1990).

The following definitions are provided in order to aid those skilled in the art in understanding the detailed description of the present invention.

5 The term “promoter” or “promoter region” refers to a nucleic acid sequence, usually found upstream (5’) to a coding sequence, that controls expression of the coding sequence by controlling production of messenger RNA (mRNA) by providing the recognition site for RNA polymerase or other factors necessary for start of transcription at the correct site. As contemplated herein, a promoter or promoter region includes variations of promoters derived by
10 means of ligation to various regulatory sequences, random or controlled mutagenesis, and addition or duplication of enhancer sequences. The promoter region disclosed herein, and biologically functional equivalents thereof, are responsible for driving the transcription of coding sequences under their control when introduced into a host as part of a suitable recombinant vector, as demonstrated by its ability to produce mRNA.

15 “Regeneration” refers to the process of growing a plant from a plant cell (e.g., plant protoplast or explant).

 “Transformation” refers to a process of introducing an exogenous nucleic acid sequence (e.g., a vector, recombinant nucleic acid molecule) into a cell or protoplast in which that exogenous nucleic acid is incorporated into a chromosome or is capable of autonomous
20 replication.

 A “transformed cell” is a cell whose DNA has been altered by the introduction of an exogenous nucleic acid molecule into that cell.

 The term “gene” refers to chromosomal DNA, plasmid DNA, cDNA, synthetic DNA, or other DNA that encodes a peptide, polypeptide, protein, or RNA molecule, and regions flanking
25 the coding sequence involved in the regulation of expression.

 “Identity” refers to the degree of similarity between two nucleic acid or protein sequences. An alignment of the two sequences is performed by a suitable computer program. A widely used and accepted computer program for performing sequence alignments is CLUSTALW v1.6 (Thompson et al., 1994). The number of matching bases or amino acids is
30 divided by the total number of bases or amino acids and multiplied by 100 to obtain a percent

identity. For example, if two 580 base pair sequences had 145 matched bases, they would be 25 percent identical. If the two compared sequences are of different lengths, the number of matches is divided by the shorter of the two lengths. For example, if there were 100 matched amino acids between 200 and a 400 amino acid proteins, they are 50 percent identical with respect to the shorter sequence. If the shorter sequence is less than 50 bases or amino acids in length, the number of matches are divided by 50 and multiplied by 100 to obtain a percent identity.

“C-terminal region” refers to the region of a peptide, polypeptide, or protein chain from the middle thereof to the end that carries the amino acid having a free carboxyl group.

The phrase “DNA segment heterologous to the promoter region” means that the coding DNA segment does not exist in nature in the same gene with the promoter to which it is now attached.

The term “encoding DNA” refers to chromosomal DNA, plasmid DNA, cDNA, or synthetic DNA that encodes any of the enzymes discussed herein.

The term “genome” as it applies to bacteria encompasses both the chromosome and plasmids within a bacterial host cell. Encoding DNAs of the present invention introduced into bacterial host cells can therefore be either chromosomally integrated or plasmid-localized. The term “genome” as it applies to plant cells encompasses not only chromosomal DNA found within the nucleus, but organelle DNA found within subcellular components of the cell. DNAs of the present invention introduced into plant cells can therefore be either chromosomally integrated or organelle-localized.

The terms “microbe” or “microorganism” refer to algae, bacteria, fungi, and protozoa.

The term “mutein” refers to a mutant form of a peptide, polypeptide, or protein.

“N-terminal region” refers to the region of a peptide, polypeptide, or protein chain from the amino acid having a free amino group to the middle of the chain.

“Overexpression” refers to the expression of a polypeptide or protein encoded by a DNA introduced into a host cell, wherein said polypeptide or protein is either not normally present in the host cell, or wherein said polypeptide or protein is present in said host cell at a higher level than that normally expressed from the endogenous gene encoding said polypeptide or protein.

The term “plastid” refers to the class of plant cell organelles that includes amyloplasts, chloroplasts, chromoplasts, elaioplasts, eoplasts, etioplasts, leucoplasts, and proplastids. These

organelles are self-replicating and contain what is commonly referred to as the “chloroplast genome,” a circular DNA molecule that ranges in size from about 120 kb to about 217 kb, depending upon the plant species, and which usually contains an inverted repeat region.

The phrase “simple carbohydrate substrate” means a monosaccharide or an
5 oligosaccharide but not a polysaccharide; simple carbohydrate substrate includes glucose, fructose, sucrose, lactose. More complex carbohydrate substrates commonly used in media such as corn syrup, starch, and molasses can be broken down to simple carbohydrate substrates.

The term “solids” refers to the nonaqueous component of a tuber (such as in potato) or a
10 fruit (such as in tomato) comprised mostly of starch and other polysaccharides, simple carbohydrates, nonstructural carbohydrates, amino acids, and other organic molecules.

The following examples are provided to better elucidate the practice of the present invention and should not be interpreted in any way to limit the scope of the present invention. Those skilled in the art will recognize that various modifications, truncations, etc., can be made to the methods and genes described herein while not departing from the spirit and scope of the
15 present invention.

EXAMPLES

EXAMPLE 1

cDNA cloning and overexpression

20 Unless otherwise stated, basic DNA manipulations and genetic techniques, such as PCR, agarose electrophoresis, restriction digests, ligations, *E. coli* transformations, colony screens, and Western blots were performed essentially by the protocols described in Sambrook et al. (1989) or Maniatis et al. (1982).

25 The *E. coli fda* gene sequence (SEQ ID NO: 1) was obtained from Genbank (Accession Number X14682) and nucleotide primers with homology to the 5' and 3' end were designed for PCR amplification. *E. coli* chromosomal DNA was extracted and the *E. coli fda* gene was amplified by PCR using the 5' oligonucleotide

5'GGGGCCATGGCTAAGATTTTGGATTTCGTA3' (SEQ ID NO:3) and the 3'

30 oligonucleotide 5'CCCCGAGCTCTTACAGAACGTCGATCGCGTTCAG3' (SEQ ID NO:4).

The PCR cycling conditions were as follows: 94°C, 5 min (1 cycle); addition of polymerase; 94°C, 1 min., 60°C, 1 min., 72°C, 2 min.30 sec. (35 cycles). The 1.08 kb PCR product was gel purified and ligated into an *E.coli* expression vector, pMON5723, to form a vector construct that was used for transformation of frozen competent *E.coli* JM101 cells. The pMON5723 vector contains the *E.coli* recA promoter and the T7 gene10 leader (G10L) sequences, which enable high level expression in *E.coli* (Wong et al., 1988). After induction of the transformed cells, a distinct protein band of about 40 kDa was apparent on an SDS PAGE gel, which correlates with the size of the subunit polypeptide chain of the dimeric aldolase II. It was shown that most of the induced protein was present in the soluble phase. A gel slice containing the highly induced protein was isolated and antibodies were produced in a goat, which was injected with the homogenized gel slice (emulsified in Freund's complete adjuvant).

The *fda* gene sequence was subsequently cloned into another *E.coli* expression vector, under the control of the *taq* promoter. Induction with IPTG of JM101 cells transformed with this vector showed the same 40 kDa overexpressed protein band. This new clone was used in an enzyme assay for FDA activity. Cells transformed with this vector construct were grown in a liquid culture, induced with IPTG, and grown for another 3 hours. Subsequently, a 3 mL cell culture was spun down, dissolved in 100mM Tris and sonicated. The cell pellet was spun down, and the crude cell extract supernatant was assayed for FDA activity, using a coupled enzymatic assay as described by Baldwin et al. (1978). This assay was routinely performed at 30°C.

The reaction was performed in a 1 mL final volume in excess presence of the enzymes triosephosphate isomerase (TIM) and alpha-glycerophosphate dehydrogenase (GDH) in a reaction mixture containing final concentrations of 100mM Tris pH 8.0, 4.75 mM fructose 1,6 bisphosphate, 0.15 mM NADH, 500 U/mL TIM, and 30 U/mL GDH.

The decrease in absorbance at 340nm, after addition of the cell extract supernatant, was recorded using an HP diode array spectrophotometer. One international unit (I.U.) of aldolase activity is that causing the oxidation of 2 µmol of NADH/min in this assay system.

Cell extracts containing the vector with the *fda* sequence showed a substantial increase in aldolase activity (13.1 I.U./mg protein) as compared to cells transformed with the control vector

(0.15 I.U./mg protein). The activity was shown to be inhibited by EDTA, known to specifically inhibit class II aldolases.

EXAMPLE 2

5 *Plant transformation and fda expression in tobacco*

Targeting of FDA protein

E.coli fructose 1,6 biphosphate aldolase was targeted to the plastid in plants in order to assess its influence on carbohydrate metabolism and starch biosynthesis in these plant organelles.

10 To accomplish the import of the *E.coli* aldolase into the plastids, a vector was constructed in which the aldolase was fused to the *Arabidopsis* small subunit transit peptide (CTP1) (Stark et al., 1992) or the maize small subunit CTP (Russell et al., 1993), creating constructs in which the CTP-*fda* fusion gene was located between the 35S promoter from the figwort mosaic virus (P-FMV35S; Gowda et al., 1989) and the 3'-nontranslated region of the nopaline synthase gene
15 (NOS 3'; Fraley et al., 1983) sequences. The vector construct containing the expression cassette [P-FMV/CTP1/*fda*/NOS3'] was subsequently used for tobacco protoplast transformation, which was performed as described in Fromm et al. (1987), with the following modifications. Tobacco cultivar Xanthi line D (Txd) cell suspensions were grown in 250-mL flasks, at 25°C and 138 rpm in the dark. For maintenance, a sub-culture volume of 9 mL was removed and added to 40 mL of
20 fresh Txd media containing MS salts, 3% sucrose, 0.2 g/L inositol, 0.13 g/L asparagine, 80 µL of a 50 mg/mL stock of PCPA, 5 µL of a 1 mg/mL stock of kinetin, and 1 mL of 1000x vitamins (1.3 g/L nicotinic acid, 0.25 g/L thiamine, 0.25 g/L pyridoxine HCL, and 0.25 g/L calcium pantothenate) every 3 to 4 days. Protoplasts were isolated from 1-day-old suspension cells that came from a 2-day-old culture. Sixteen milliliters of cells were added to 40 mL of fresh Txd
25 media and allowed to grow 24 hours prior to digestion and isolation of the protoplasts. The centrifugation stage for the enzyme mix has been eliminated. The electroporation buffer and protoplast isolation media were filter sterilized rather than autoclaved. The electroporation buffer did not have 4 mM CaCl₂ added. The suspension cells were digested in enzyme for 1 hour. Protoplasts were counted on a hemacytometer, counting only the protoplasts that look

intact and circular. Bio-rad Gene Pulser cuvettes (catalog # 165-2088) with a 0.4-cm gap and a maximum volume of 0.8 mL were used for the electroporations. Fifty to 100 µg of DNA containing the gene of interest along with 5 µg of internal control DNA containing the luciferase gene were added per cuvette. The final protoplast density at electroporation was 2×10^6 /mL and electroporator settings were a 500 µFarad capacitance and 140 volts on the Bio-rad Gene Pulser. Protoplasts were put on ice after resuspension in electroporation buffer and remained on ice in cuvettes until 10 minutes after electroporation. Protoplasts were added to 7 mL of Txd media + 0.4 M mannitol and conditioning media after electroporation. At this stage coconut water was no longer used. The protoplasts were grown in 1- hour day/night photoperiod regime at 26°C and were spun down and assayed or frozen 20-24 hours after electroporation.

Western blot analysis performed on the protoplast extracts, obtained after transformation, showed processing into the mature FDA in the tobacco protoplasts. Expression was detected of a protein migrating at approximately 40 kDa, which is the molecular weight of the aldolase subunit and the size of the protein also observed after overexpression of the aldolase in *E. coli*.

The expression cassette [P-FMV/CTP1/*fda*/NOS3'] was subsequently cloned into the NotI site of pMON17227 (described in PCT Publication WO 92/04449), in the same orientation as the selectable marker expression cassette, to form the plant transformation vector pMON17524, as shown in Figure 2 (SEQ ID NO: 5).

An additional construct was made and used for tobacco protoplast transformation, fusing the *fda* gene to the *Arabidopsis* EPSPS transit peptide (CTP2), which is described in US patent 5,463,175. The transit peptide was cloned (through the SphI site) into the SphI site located immediately upstream from the N-terminus of the *fda* gene sequence in the CTP1-*fda* fusion (described above). This new CTP2-*fda* fusion gene was then cloned into a vector between the FMV promoter and the NOS 3' sequences. When this construct containing the CTP2/*fda* gene sequences was used for tobacco protoplast transformation, expression was detected of a protein migrating at approximately 40 kDa, which is the molecular weight of the aldolase subunit and the size of the protein also observed after overexpression of the aldolase in *E. coli*.

The NotI cassette [P-FMV/CTP2/*fda*/NOS3'] from this construct was then cloned into the NotI site of pMON17227, in the same orientation as the selectable marker expression cassette, to form the plant transformation vector pMON17542, which is shown in Figure 3 (SEQ ID NO:6).

For cytoplasmic expression of the FDA in tobacco protoplasts, a construct was made in which the *fda* gene sequence (without being coupled to a transit peptide) was cloned into a vector backbone, between the FMV promoter and the NOS 3' sequences. Using this construct for tobacco protoplast transformation also showed expression of a protein of the same size, migrating at approximately 40 kDa.

fda expression in tobacco plants

Two constructs, containing the *fda* gene, fused to the *Arabidopsis* small subunit CTP1 (pMON17524) (SEQ ID NO:5, Figure 2) and the *Arabidopsis* EPSPS (CTP2) transit peptide (pMON17542) (SEQ ID NO:6, Figure 3), were used for tobacco plant transformation, as described in US patent 5,463,175. A vector without the CTP-*fda* sequences, pMON17227 (described in PCT Publication WO 92/04449), was used as a negative control. The plant transformation vectors were mobilized into the ABI *Agrobacterium* strain. Mating of the plant vector into the ABI strain was done by the triparental conjugation system using the helper plasmid pRK2013 (Ditta et al., 1980).

Growth chamber-grown tobacco transformant lines were generated and first screened by Western blot analysis to identify expressors using goat antibody raised against *E.coli*-expressed *fda*. Subsequently, for pMON17524-expressing tobacco lines, leaf nonstructural carbohydrates were analyzed (sucrose, glucose, and hydrolyzed starch into glucose) by means of a YSI Instrument, Model 2700 Select Biochemistry Analyzer. Starting at flowering stage, leaf samples were also taken from these plants and analyzed for diurnal changes in leaf nonstructural carbohydrates.

Five hundred milligrams to 1 g fresh tobacco leaf tissue samples were harvested and extracted in 5 mL of hot Na-phosphate buffer (40 g/L NaH₂PO₄ and 10 g/L Na₂H₂PO₄ in double de-ionized water) by homogenization with a Polytron. Test tubes were then placed in an 85°C water bath for 15 minutes. Tubes were centrifuged for 12 minutes at 3000 rpm and the supernatants saved for soluble sugar analysis. The pellet was resuspended in 5 mL of hot Na-

phosphate buffer mixed with a Vortex and centrifuged as described above. The supernatant was carefully removed and added to the previous supernatant fraction for soluble sugar (sucrose and glucose) analysis by YSI using appropriate membranes.

The starch was extracted from the pellet using the Megazyme Kit (Megazyme, Australia).

- 5 To the pellet, 200 μ L of 50% ethanol and 3 mL of thermostable alpha-amylase (300U) were added and the mixture vortexed. Samples were then incubated in boiling water for 6 minutes and stirred after 2 and 4 minutes. Tubes were placed in 50°C water bath and 4 mL of 200 mM acetate buffer (pH 4.5) were added followed by 0.1 mL amyloglucosidase (20 U). Incubation occurred for 1 hour. Test tubes were then centrifuged for 15 minutes at 3000 rpm. Aliquots were taken
- 10 from the supernatant and analyzed for glucose by YSI. The free glucose was adjusted to anhydrous glucose (as it occurs in starch by multiplying by the ratio 162/182). The total volume per tube was 7.1 mL.

- As seen in Table1, expression of the *fda* gene in tobacco correlated with a significant increase in leaf starch levels. However, referring to Figure 4, when a diurnal profile of starch
- 15 levels was established in the *fda*-expressing leaves, this increase was apparent mainly early in the photoperiod, which is a phase when leaves are known to have peak photosynthetic activity. This increase in starch has no apparent negative effect on the plant because the increased starch is turned over during the dark period. There was no apparent increase in steady state levels of sucrose or glucose in tobacco leaves expressing *E.coli fda* as compared to the control.

20

Table 1
Leaf Carbohydrate Levels of Plants Expressing
the *fda* Transgene¹ (pMON17524)

	High Expressors (>0.01% total protein)	Low Expressors (< 0.01%) (mg/g fresh weight)	Negative Control
5			
	STARCH 35.08 ± 2.84	23.25 ± 3.20	16.69 ± 2.92
10	SUCROSE 0.97 ± 0.17	0.86 ± 0.25	0.66 ± 0.19
	GLUCOSE 1.88 ± 0.17	1.58 ± 0.20	1.68 ± 0.26

¹ Leaf samples were harvested at midday.

A second set of transgenic tobacco plants transformed with the construct pMON17542 were grown in the greenhouse. Tobacco plants containing a vector without the CTP-*fda* sequences, pMON17227, were used as negative control. Of all the pMON17542-lines screened for expression by Western blot analysis, 18 were high expressors (>0.01% of the total cellular protein) and 15 lines were low expressors (<0.01%). Fifteen plants containing the null vector, pMON17227, were used as control. Fully expanded leaves from plants expressing the *fda* transgene and negative controls were tested for sucrose export by collecting phloem exudate from excised leaf systems. The phloem exudation technique is described in Groussol et al. (1986). Leaves were harvested at 11:30 AM and placed in an exudation medium, containing 5 mM EDTA at pH 6.0, and allowed to exude for a period of 4 hours under full light and high humidity. The exudation solution was immediately analyzed for sucrose level, as described above in the carbohydrate analysis method. As seen in Table 2, a significant increase in sucrose export out of source leaves was observed in plants expressing the *fda* transgene.

This increase in sucrose export by *fda*-expressing leaves is an illustration of an increase in source capacity, very likely due to an increased carbon flow through the Calvin Cycle (in response to increased triose-P utilization) and thus an increase in net carbon utilization by the

leaf. As seen in Table 2, the increase in sucrose loading in the phloem correlates with the level of *fda* expression.

Table 2

Levels of Sucrose in Phloem Exudate from
Excised Leaves of *fda* Transgenic Tobacco Plants (pMON17542)

	Water uptake ($\mu\text{l/g F.Wt./h}$)	sucrose in phloem exudate (ng/leaf)	(ng/g F.Wt.)
<i>fda</i> high expressors	320 ± 20	330 ± 60	108 ± 22
<i>fda</i> low expressors	340 ± 10	210 ± 10	77 ± 3
Control	390 ± 30	160 ± 10	56 ± 3

Referring to Table 3, preliminary analysis of plant growth and development revealed no significant differences in number of leaves or pods per plant, plant height, stem diameter, or apparent seed weight per plant, between plants expressing the *fda* gene and the vector control under the specific growing and analysis conditions. However, as seen in Table 4, the *fda*-transgenic plants had a significantly higher root mass. This may be an indication that, under these conditions, roots represented a more dominant sink that attracted excess carbohydrate produced by the source leaves. Furthermore, the present illustration shows that the increase in root mass in the presence of the *E.coli fda* gene was accomplished with no apparent negative effect on shoot growth, inflorescence, or seed set. Therefore, this increase in root growth and final root dry weight is a desirable plant trait because it would lead to a rapid seedling establishment following germination and greater plant ability to tolerate drought, cold stress, other environmental challenges, and transplanting. In different plants and under different growing conditions, other plant parts (such as seed, fruit, stem, leaf, tuber, bulb, etc.) are expected to show the weight increase observed in tobacco roots overexpressing the *fda* transgene.

Table 3
Assessment of Certain Plant Growth and Development Parameters in
Tobacco Expressing the *fda* Transgene¹ (pMON17542)

	#pods/plant	#leaves/plant	Plant height (cm)	Seed weight (g/plant)
high expressors	162 ± 40	25.4 ± 0.8	65.3 ± 3.1	18.8 ± 2.4
Control	156 ± 28	24.4 ± 0.5	65.8 ± 5.1	17.3 ± 2.6

¹ To achieve this analysis, 14 high-expressor lines were compared to 15 control plants. Measurements were made prior to seed harvest (most pods have reached maturity). The number of leaves was confirmed by counting the number of nodes to account for leaf drop.

Table 4
Tobacco Root Dry Weight of Plants Expressing
the *E.coli fda* Transgene¹ (pMON17542)

	Root Dry Weight (g/plant)
<i>fda</i> high expressors	64.0 ± 3.9
<i>fda</i> low expressors	62.7 ± 5.4
Control	31.7 ± 1.6

¹ Roots from 5 high and 7 low expressing lines and 6 control plants were excised and washed carefully then placed in a 65°C drying oven for at least 48 hours. Roots were removed from the oven and allowed to equilibrate in the laboratory for 2 hours before dry weight determination.

EXAMPLE 3

Plant transformation and fda expression in corn plants

Targeting of FDA protein

5 Vectors containing the *fda* gene with and without the plastid targeting peptide were made for transformation in corn and are also suitable for other monocots, including rice, wheat, barley, sugarcane, triticale, etc.

 For the cytosolic expression of the *fda* gene in corn plants, a construct was made in which the *fda* gene sequence was fused to the backbone of a vector containing the enhanced
10 CaMV 35S promoter (e35S; Kay et al., 1987), the HSP70 intron (US patent 5,593,874), and the NOS3' polyadenylation sequence (Fraley et al., 1983). This created a NotI cassette [P-e35S/HSP70 intron/*fda*/NOS3'] that was cloned into the NotI site of pMON30460, a monocot transformation vector, to form the plant transformation vector pMON13925, as shown in Figure 5. pMON30460 contains an expression cassette for the selectable marker neomycin
15 phosphotransferase typeII gene (nptII) [P-35S/NPTII /NOS3'] and a unique NotI site for cloning the gene of interest. The final vector (pMON13925) was constructed so that the gene of interest and the selectable marker gene were cloned in the same orientation. A vector fragment containing the expression cassettes for these gene sequences could be excised from the bacterial selector (Kan) and ori, gel purified, and used for plant transformation.

20 For the chloroplast-targeted expression of the *fda* gene in corn plants, a construct was made in which the *fda* gene sequence, coupled to the maize RUBISCO small subunit CTP (Russell et al., 1993), was fused to the backbone of a vector containing the enhanced (CaMV) 35S promoter, the HSP70 intron, and the NOS3' polyadenylation sequences. This created a NotI cassette [P-e35S/HSP70 intron/mzSSuCTP/*fda*/NOS3'] that was cloned into the NotI site (in the
25 same orientation as the selectable marker cassette [P-35S/NPTII /NOS3']) of the monocot transformation vector pMON30460, to form the vector pMON17590, as shown in Figure 6. From this vector a fragment containing the *fda* gene expression cassette and the selectable marker cassette could be excised from the bacterial selector (Kan) and ori, gel purified, and used for plant transformation.

For the cytosolic endosperm-specific expression of the aldolase gene in corn, the *fda* gene sequence was cloned into a vector (in the same orientation as the selectable marker cassette [P-35S/NPTII /NOS3']) containing the glutelin gene promoter P-osgt1 (Zheng et al., 1993), the HSP70 intron, and the NOS3' polyadenylation sequences to form the vector pMON13936, as shown in Figure 7. From this vector a fragment containing the *fda* gene expression cassette [P-osgt1/HSP70intron/*fda*/NOS3'] and the selectable marker cassette could be excised from the bacterial selector (Kan) and ori, gel purified, and used for plant transformation.

Maize plant transformation

Transgenic maize plants transformed with the vectors pMON13925 (described above) or pMON17590 (described above) were produced using microprojectile bombardment, a procedure well-known to the art (Fromm, 1990; Gordon-Kamm et al., 1990; Walters et al., 1992).

Embryogenic callus initiated from immature maize embryos was used as a target tissue. Plasmid DNA at 1mg/mL in TE buffer was precipitated onto M10 tungsten particles using a calcium chloride / spermidine procedure, essentially as described by Klein et al. (1988). In addition to the gene of interest, the plasmids also contained the neomycin phosphotransferase II gene (*nptII*) driven by the 35S promoter from Cauliflower Mosaic Virus. The embryogenic callus target tissue was pretreated on culture medium osmotically buffered with 0.2M mannitol plus 0.2M sorbitol for approximately four hours prior to bombardment (Vain et al., 1993). Tissue was bombarded two times with the DNA-coated tungsten particles using the gunpowder version of the BioRad Particle Delivery System (PDS) 1000 device. Approximately 16 hours following bombardment, the tissue was subcultured onto a medium of the same composition except that it contained no mannitol or sorbitol, and it contained an appropriate aminoglycoside antibiotic, such as G418", to select for those cells that contained and expressed the 35S/*nptII* gene. Actively growing tissue sectors were transferred to fresh selective medium approximately every 3 weeks. About 3 months after bombardment, plants were regenerated from surviving embryogenic callus essentially as described by Duncan and Widholm (1988).

Aldolase activity from transgenic maize

In order to measure leaf aldolase activity, corn leaf samples were taken and immediately frozen on dry ice. Aldolase enzyme was extracted from the leaf tissue by grinding the leaf tissue

at 4°C in 1.2 mL of the extraction buffer (100 mM Hepes, pH 8.0, 5 mM MgCl₂, 5 mM MnCl₂, 100 mM KCl, 10 mM DTT, 1% BSA, 1 mM PMSF, 10 µg/mL leupeptin, 10 µg/mL aprotinin). The extract was centrifuged at 15,000 x g, at 4°C for 3 minutes, and the non-desalted supernatant was assayed for enzyme activity. This extraction method gave about 60% recovery of *E. coli*

5 FDA activity.

Total aldolase activity was determined in 0.98 mL of reaction mixture that consisted of 100 mM EPPS-NaOH, pH 8.5, 1 mM fructose-bisphosphate, 0.1 mM NADH, 5 mM MgCl₂, 4 units of alpha-glycerophosphate dehydrogenase, and 15 units of triosephosphate isomerase. The reaction was initiated by addition of 20 µL of leaf extract. The resulting data, generated from a single experiment, are presented in Table 5.

Table 5
Aldolase Activity from Transgenic Maize Leaves

Lines	A340/min/20µL	Activity %
H99 (control)	0.113	100
pMON 17590	0.233	206
pMON13925	0.251	222

20 A phenotype was visible in the primary transformants (RO plants) expressing the *E. coli* FDA when the protein was targeted to the chloroplast. The leaves were chlorotic but seed set was normal. R1 plants were grown in both field and in greenhouse experiments. Starch was not detectable in the leaves using an iodine staining and pollination was delayed. It is believed that the phenotype in these corn plants may be the result of the promoter (e35S) used in both the pMON17590 and pMON13925 vectors not being preferred for causing FDA expression in corn. Because e35S is believed to cause mesophyll enhanced expression and the Calvin Cycle in a C4 plant such as corn occurs predominantly in the bundle sheath cells, the use of a promoter directing enhanced expression in the bundle sheath cells (such as the ssRUBISCO promoter)

may be preferred. Vectors containing such a promoter and driving expression of FDA have been prepared and are being tested in maize.

In particular, the maize RuBISCO small subunit (PmzSSU, a bundle sheath cell-specific promoter) has been used to construct vectors for cell-specific *fda* expression in maize. A class I aldolase (*fdaI*), an *fda* without an iron sulfur cluster and with different properties from *fdaII*, was utilized to improve carbon metabolism in C4 crops (e.g. maize) . The gene for the class I aldolase was amplified from the genome of *Staphylococcus aureus* and activity was confirmed. Transformation vectors were then constructed to express both classes of aldolase (*fdaI* and *fdaII*) in a cell-specific manner in maize. The following cassettes have been made:

pMON13899: PmzSSU/hsp70/mzSSU CTP/*fdaI*

pMON13990PmzSSU/hsp70/mzSSU CTP/*fdaII*

pMON13988:P35S/hsp70/*fdaI*.

These vectors were used for corn transformation as described generally above. The biochemical and physiological analysis of the primary transformants should allow for the identification of aldolase gene overexpression that will lead to increase growth and development and yield in maize.

Also, two vectors were used for transformation of corn which would target the expression of the *E. coli fda II* gene in the maize endosperm. The vector pMON 13936 uses the rice *gt1* promoter to drive expression of aldolase in the cytoplasm of the endosperm cells. Another vector (pMON 36416) uses the same promoter with the maize RuBISCO small subunit transit peptide to localize the protein in the amyloplasts. Homozygous lines of the cytosolic aldolase transformants have been identified (Homozygosity of 37 plants was confirmed using western blot analysis) and seed from these plants were collected for grain composition analysis (moisture, protein, starch, and oil). Of the 53 pMON 36416 primary transformants screened for amyloplast-targeted aldolase expression, 11 were positive. These plants will be tested for homozygosity selection/propagation and kernels from the homozygotes will be used for composition analysis.

EXAMPLE 4

Plant transformation and fda expression in potato plants

Targeting of fda expression

5 The plant expression vector, pMON17542 (described earlier), in which the *fda* gene is expressed behind the FMV promoter and the aldolase enzyme is fused to the chloroplast transit peptide CTP2, was used for *Agrobacterium*-mediated potato transformation.

 A second potato transformation vector was constructed by cloning the NotI cassette [P-FMV/CTP2/*fda*/NOS3'] (described earlier) into the unique NotI site of pMON23616.

10 pMON23616 is a potato transformation vector containing the nopaline-type T-DNA right border region (Fraley et al., 1985), an expression cassette for the neomycin phosphotransferase typeII gene [P-35S/NPTII /NOS3'] (selectable marker), a unique NotI site for cloning the gene expression cassette of interest, and the T-DNA left border region (Barker et al., 1983). Cloning of the NotI cassette [P-FMV/CTP2/*fda*/NOS3'] (described earlier) into the NotI site of
15 pMON23616 results in the potato transformation vector pMON17581, as shown in Figure 8. The vector pMON17581 was constructed such that the gene of interest and the selectable marker gene were transcribed in the same direction.

Potato plant transformation

20 The plant transformation vectors were mobilized into the ABI *Agrobacterium* strain. Mating of the plant vector into the ABI strain was done by the triparental conjugation system using the helper plasmid pRK2013 (Ditta et al., 1980). The vector pMON17542 was used for potato transformation via *Agrobacterium* transformation of Russet Burbank potato callus, following the method described in PCT Publication WO 96/03513 for glyphosate selection of
25 transformed lines.

 After transformation with the vector pMON17542, transgenic potato plantlets that came through selection on glyphosate were screened for expression of *E. coli* aldolase by leaf Western blot analysis. Out of 112 independent lines assayed, 50 *fda*-expressing lines (45%) were identified, with *fda* expression levels ranging between 0.12% and 1.2 % of total extractable
30 protein.

The plant transformation vector PMON17581 was used for *Agrobacterium*-mediated transformation of HS31-638 potato callus. HS31-638 is a Russet Burbank potato line previously transformed with the mutant ADPglucose pyrophosphorylase (*glgC16*) gene from *E.coli* (U.S. Patent 5,498,830). The potato callus was transformed following the method described in PCT
5 Publication WO 96/03513, substituting kanamycin (administered at a concentration of 150-200 mg/L) for glyphosate as a selective agent.

The transgenic potato plants were screened for expression of the *fda* gene by assaying leaf punches from tissue culture plantlets. Western blot analysis (using antibodies raised against the *E. coli* aldolase) of leaf tissue from the pMON17581-transformed lines identified 12
10 expressing lines out of 56 lines screened. Expression was detected of a protein migrating at approximately 40 kDa, which is the molecular weight of the *E. coli* (classII) aldolase subunit and the size of the protein observed after overexpression of the aldolase in *E. coli*.

Specific gravity measurements of transgenic potato plants

15 From the 50 *fda*-expressing potato lines obtained after transformation with pMON17542, 7 of the highest expressing lines were micropropagated in tissue culture, and 8 copies of each line were planted in pots 14 inches in diameter and 12 inches deep, containing a mixture of: ½ Metro 350 potting media, ¼ fine sand, ¼ Ready Earth potting media. Wild-type Russet Burbank plantlets from tissue culture were planted as controls. All plants were cultivated for
20 approximately 5 months in the greenhouse in which daytime temperature was approximately 21-23°C while nighttime temperature was approximately 13°C. Plants were watered every other day throughout their active growing period and fertilized with Peter's 20-20-20 commercial fertilizer once a week, at levels similar to commercial applications. Fertilization was carried out only for the first 2 ½ months, at which point fertilization was stopped completely. Plants were allowed to
25 naturally senesce, and at approximately 50% senescence, tubers were harvested.

For each line at harvest, all tubers from all 8 pots were pooled and a total weight was obtained. Then for each line, tubers 30 g or greater were pooled and specific gravity was determined on this group of tubers. Specific gravity is the weight of the tubers in air divided by the weight in air minus the weight in water. Results of these weight measurements are presented
30 in Table 6.

Table 6
Specific gravity measurements from transgenic potato plants

Line #	Total Weight	Overall % Yield Increase	Combined Weight of Tubers over 30g	% Increase in Total Weight (Tubers over 30g)	Combined Weight of Tubers over 30g (% of Total Weight)	Specific Gravity
RB	6609		4477		67.70%	1.087
40652	5138	neg	1307	neg	25.40%	1.08
40611	7170	8.5%	4533	1.3%	63.20%	1.083
40608	7470	13.0%	1070	neg	14.30%	1.081
40632	7776	21.8%	5453	21.8%	70.10%	1.088
40614	8688	31.5%	5468	22.2%	62.90%	1.083
40631	8800	33.2%	6188	38.2%	70.30%	1.084
40610	9746	47.0%	7777	73.0%	80%	1.087

This table summarizes the tuber yield and specific gravity for all seven lines grown in the greenhouse. The results indicate that, in comparison to the control, all but one of the *fda* lines show an increase in overall tuber yield, and that in four lines, there is a corresponding increase in percentage of tubers that weigh more than 30 g. For combined tubers over 30 g, the percent of total weight is near that of the control, and for two lines is greater than the control. This indicates that five out of the six of the lines show higher overall yield and are not making smaller tubers. In other words, with the increase in overall yield, there is a corresponding increase in percentage of bigger tubers (over 30 g). However, there is no increase in specific gravity of the tubers.

In conclusion, it appears that expression of *fda* in potato produces greater numbers of tubers per plant without a sacrifice in tuber size. This represents a yield benefit in that the farmer could potentially be able to produce the same amount of tubers using less acreage. Similar experiments will also be performed by co-expression of *fda* with other carbohydrate metabolizing genes, such as *glgC16*, in order to determine how such combinations will affect tuber yield, tuber solids deposition and overall tuber specific gravity.

Aldolase activity from transgenic potato

After being cultivated for 3 months (post planting) in the greenhouse, leaf samples were taken from 6 of the highest *fda*-expressing potato lines, obtained after transformation with pMON17542, and assayed for aldolase activity.

In order to measure potato leaf aldolase activity, duplicate leaf samples from each line were taken and immediately frozen on dry ice. Aldolase was extracted from 0.2 g of leaf tissue by grinding at 4°C in 1.2 mL of the extraction buffer: 100 mM Hepes, pH 8.0, 5 mM MgCl₂, 5 mM MnCl₂, 100 mM KCl, 10 mM DTT, 1% BSA, 1mM PMSF, 10 µg/mL leupeptin, 10 µg/mL aprotinin. The extract was assayed for aldolase activity as described earlier.

Six independent transgenic potato lines expressing *fda* were tested for aldolase activity. The expression of *fda* in leaves is an indicator of the expression in the whole plant because the FMV promoter used to drive expression of the respective encoding DNAs directs gene expression constitutively in most, if not all, tissues of potato plants.

Table 7 summarizes the quantitative protein expression data for each of the lines, and the percent activity for each individual line.

Table 7
Aldolase Activity from
Transgenic Russet Burbank Potato Leaves

Lines	Exp. #1		Exp. #2		Average % Activity
	Act (U/gFW)	%Act	Act (U/gFW)	%Act	
Control	4.461	100	4.732	100	100
40608	6.969	156	8.055	170	163
40610	8.489	190	7.326	155	173
40652	5.812	130	6.367	135	132
40632	5.257	118	4.244	90	104
40631	5.764	129	4.968	105	117
40611	5.715	128	5.836	123	126

Solids uniformity in transgenic potato

Twenty-five Russet Burbank lines expressing *fda* (potato lines designated "Maestro"), obtained after transformation with pMON17542, and fifteen Russet Burbank Simple Solid lines, also containing *glgC16* (PCT Publication WO 91/19806 and US Patent 5,498,830), expressing *fda* (potato lines designated "Segal"), obtained after transformation with pMON17581, were field tested at two different sites. For each field site, 36 plants per line (three repetitions of 12 plants per line) were evaluated for tuber solids distribution. At harvest, tubers were pre-sorted at each field site into a ten to twelve ounce category, and nine tubers from each replicated plot were analyzed in groups of three.

For a typical 10-12 ounce tuber having a diameter of 7-8 cm, starch distribution was evaluated by removing the center longitudinal slice (13 mm) from each tuber. Slices were then peeled and laid flat on a cutting board where the inner tuber region (pith region) was removed by a 14-mm cork punch. The tissue from pith to cortex (perimedullary region) was removed by an up-to-a 2-inch cork punch. The remaining cortex tissue was approximately an 8-mm wide ring from the outermost region of the slice.

Specific gravity was then determined by weighing both the pooled pith punches and pooled cortex punches in air and then in water:

$$\text{Specific gravity} = \text{Air Wt.}/(\text{Air Wt.}-\text{Water Wt.})$$

After calculating specific gravity, solids levels were determined by the following equation:

$$-214.9206 + (218.1852 * \text{Sp. Gravity})$$

The degree of solids uniformity (Solids Uniformity Index) is determined by calculating the pith to cortex solids ratio (pith solids divided by cortex solids). The three groups of three tubers per plot were averaged, at which point the average of three plot replications was calculated per field site.

Analyses of several previous solids uniformity field trials (data not shown) have demonstrated nontransgenic, wild-type Russet Burbank potato to have a typical pith to cortex tuber solids ratio within the range of 68% to 72%, depending on growing region and agricultural practices. Tables 8-11 provide the pith to cortex solids ratios by plant line number, with a higher pith to cortex solids ratio indicating a greater degree of solids uniformity.

Tables 8 and 9 represent the data from one field site (site 1) for Segal and Maestro, respectively, and illustrate that the majority of Segal and Maestro lines have higher pith to cortex solids ratios than that of 68.4% for the Russet Burbank control, with some lines approaching an 82% pith to cortex solids ratio.

Tables 10 and 11 represent the data from another field site (site 2) for Segal and Maestro, respectively, and also illustrate that the majority of Maestro and Segal lines have higher pith to cortex solids ratios than that of the Russet Burbank control, with some lines approaching an 88% pith to cortex solids ratio. In the site 2 field trial, the Russet Burbank control had an atypical, abnormally high pith-to-cortex solids uniformity ratio of 79.3%, which was most likely due to environmental growing conditions. The site 2 results demonstrate that expression in Russet Burbank potato of *E. coli fda*, alone or with co-expression of *glgC16*, increases tuber solids uniformity even in a growing season when tuber solids uniformity is already extremely high in nontransgenic Russet Burbank. That is, the *fda* gene continues to perform when agricultural conditions are already conducive to an abnormally high solids uniformity level.

Table 8. Solids Uniformity Index: Pith Solids to Cortex Solids Ratio. Segal Russet Burbank Lines. Site 1

<u>Line</u>	<u>Ratio</u>
S-29	79.1
S-9	75.8
S-20	71.3
S-15	71.3
S-21	70.5
S-5	70.2
S-18	70.0
RB control	68.4
S-32	68.3
S-16	65.6

**Table 9. Solids Uniformity Index: Pith Solids to Cortex Solids Ratio.
Maestro Russet Burbank Lines. Site 1**

	<u>Line</u>	<u>Ratio</u>
5	M-13	74.0
	M-12	73.6
	M-1	73.4
	M-3	73.0
	M-6	72.4
10	M-9	71.2
	M-11	70.6
	M-18	70.5
	M-17	69.9
	M-19	69.4
15	M-5	69.3
	M-20	68.9
	RB control	68.4
	M-8	68.3
	M-43	67.7
20	M-23	67.3
	M-7	67.0
	M-39	66.6
	M-22	66.0
	M-10	65.4
25	M-27	61.4

**Table 10. Solids Uniformity Index: Pith Solids to Cortex Solids Ratio
Segal Russet Burbank Lines. Site 2**

	<u>Line</u>	<u>Ratio</u>
30	S-33	87.4
	S-54	87.1
	S-05	86.8
	S-29	85.1
	S-21	84.3
35	S-16	83.2
	S-20	81.5
	S-18	80.7
	S-32	80.6
	RB control	79.3
40	S-09	79.0

**Table 11. Solids Uniformity Index: Pith Solids to Cortex Solids Ratio
Maestro Russet Burbank Lines. Site 2**

	<u>Line</u>	<u>Ratio</u>
	M-04	87.7
5	M-18	83.9
	M-17	83.8
	M-03	83.7
	M-09	83.4
	M-15	83.2
10	M-29	82.9
	M-44	82.3
	M-08	82.2
	M-43	81.6
	M-22	81.1
15	M-05	80.8
	M-01	80.5
	M-20	80.2
	M-45	79.6
	M-39	79.5
20	M-27	79.5
	RB control	79.3
	M-13	78.9
	M-22	78.8
	M-19	78.7
25	M-07	78.2
	M-12	77.9
	M-23	77.3
	M-06	76.5
	M-10	75.0
30	M-11	74.1

The effect of aldolase on pith to cortex solids ratios in the Segal lines is slightly more dramatic than in Maestro lines. We believe this phenotype is due to expression of *fda* in a background in which the Russet Burbank host expresses *glgC16* at a relatively low to moderate level, and that the combination of *fda* plus *glgC16* provides improved benefits. Cross sectional tuber slices (Figure 9) of three Segal lines with improved solids uniformity illustrate a greater deposition of starch within the inner regions of the tuber. Specifically, an increase in cortex volume accompanied by relocation of the xylem ring towards the center of the tuber, plus a more opaque pith tissue due to an increase in starch density, are evident in the transgenic lines. This

physiological alteration may be due to an increase in sucrose translocation from source to sink, which may influence phloem element distribution during tuber development or sucrose availability for starch biosynthesis across the tuber.

5 **Example 5**

Plant transformation and FDA expression in cotton plants

The *E. coli fda* vectors pMON17524 [FMV/CTP1/*fda*] (Figure 2) and pMON17542 [FMV/CTP2/*fda*] (Figure 3) were transformed into cotton using *Agrobacterium* as described by Umbeck et al. (1987) and in US Patent 5004863. The protein was targeted to the chloroplast
10 using either the *Arabidopsis* SSU CTP 1 (pMON17524) or the *Arabidopsis* EPSPS (pMON17542) chloroplast transit peptide.

Aldolase expression in cotton

Five-week-old calli transformed with both vectors were analyzed by Western blot
15 analyses and by aldolase assays. Western blot analysis indicated a large amount of protein at the position of the full-length FDA standard and a lesser amount at the same position in the control callus extracts. It appeared that the protein was fully processed. To verify that FDA was expressed in the tissue and for comparison of activity, calli transformed with the two vectors were extracted in a buffer that would prevent loss of activity of the transgene product. BSA was
20 added to final concentration of 1 mg/mL, which limited the analysis of processing on import by Western blot. Aldolase assays were performed plus or minus 25 mM EDTA, which inhibits the *E. coli* enzyme but not the plant enzyme. The results of the assays are shown in Table 12.

Table 12

Aldolase Activity in Cotton Calli and Cotton Leaf

 $\Delta A_{340} \text{ e}^{-3}/\text{mg protein}/5 \text{ min}$

	Colony#	-EDTA	+EDTA	Fold Increase
5	<u>Controls</u>			
	Cotton Leaf (Coker)	4.0	4.2	-
	Uninoculated Calli	7.7	5.6	1.3X
	Inoculated Calli (E35S/GUS) #1	6.8	6.1	-
	#2	3.5	4.0	-
10	<u>FDA calli</u>			
	pMON 17542 #1	3.5	2.3	1.5X
	#3	5.5	2.6	2.1X
	#5	9.2	3.8	2.4X
	#4	19.8	3.6	5.5X
15	pMON17524 #2	15.2	5.8	2.6X
	#3	12.5	4.0	3.1X
	#5	14.4	2.9	4.9X
	#6	4.1	1.2	3.5X

- 20 The results indicate that there is good expression of the *fda* gene in cotton callus. Almost all calli had at least twofold higher aldolase activity, and the increase was sensitive to inhibition by EDTA. Processing appeared complete by Western blot analysis using these samples.

25

REFERENCES CITED

- Alefounder et al. (1989) Biochem. J. 257:529-534
- Ammirato et al. (1984) Handbook of Plant Cell Culture - Crop Species. Macmillan Publ. Co..
- 5 Bai et al. (1975) Arch. Biochem. Biophys. 168: 230-234.
- Baldwin et al. (1978) Biochem. J. 169: 633-641
- Barker et al. (1983) Plant Mol Biol 2 (6): 335-350.
- Benfey et al. (1989) EMBO J. 5: 2195-2202.
- Besmond et al. (1983) Biochem. Biophys. Res. Commun. 117, 601-609.
- 10 Bevan (1984) Nucleic Acids Res. 12 (22): 8711-8721.
- Bevan et al. (1986) Nucleic Acids Res. 14 (11):4625-4638.
- Campbell et al. (1994) Canadian Journal of Forest Research 24 (8):1689-1693.
- Cerdan et al. (1997) Plant Molecular Biology 33 (2): p245-255.
- Chopra et al. (1990) Plant Molecular Biology 15:517-520.
- 15 Clayton (1985) EMBO J. 4, 2997-3003.
- Cremona (1968) G. Bot. Ital. 102, 253-259.
- Daniell et al. (1998) Nature Biotechnology 16:345-348.
- Datta et al. (1990) Bio-technology 8:736-740.
- Ditta et al. (1980) Proc Natl Acad Sci USA 77(12): 7347-7351.
- 20 Duncan and Widholm (1988) Plant Cell Reports 7: 452-455.
- Edwards et al. (1990). Proc Natl Acad Sci USA 87 (9): p3459-3463.
- Fejes et al. (1990). Plant Mol Biol 15 (6): p921-932.
- Fraley et al. (1983) Proc Natl Acad Sci USA 80: 4803-4807.
- Fraley et al. (1985) Bio/Technology 3 (7): 629-635.
- 25 Fromm, M., (1990) UCLA Symposium on Molecular Strategies for Crop Improvement, April 16-22, 1990. Keystone, CO.
- Fromm et al. (1987) Methods in Enzymology. 153:351-366.
- Gordon-Kamm et al. (1990) Plant Cell 2: 603-618.
- Gotz et al. (1979) FEMS Microbiol. Lett. 5:253-257.
- 30 Gowda et al. (1989). Journal of Cellular Biochemistry supplement 13D, 301 (Abstract).

- Groussol et al. (1986) Physiologie Vegetale 24(1):123-134.
- Guerrini et al. (1971) Arch. Biochem. Biophys. 146, 249-255.
- Hannapel (1990) Plant Physiol. 94: 919-925.
- Hayashimoto et al. (1990) Plant Physiol. 93:857-863.
- 5 Herrera-Estrella et al. (1983) Nature 303:209.
- Hinchee et al., Bio/Technology 6:915-922 (1988).
- Horsch and Klee. (1986) Proc. Natl. Acad. Sci. U.S.A. 83:4428-4432.
- Jack (1973) Ph.D. Dissertation, University of Cambridge.
- Jack and Harris (1971) Biochem. J. 124, 680-690.
- 10 Jefferson et al. (1990) Plant Mol. Biol. 14: 995-1006.
- Joh et al. (1986) J. Mol. Biol. 190:401-410.
- Kay et al. (1987) Science 236: 1299-1302.
- Klee et al. (1985) Bio-Technology 3(7): 637-642.
- Klein et al. (1988) Bio/Technology 6: 559-563.
- 15 Kretsch et al. (1995) Plant Journal 7 (5): p715-729.
- Lai et al., (1974) Science 183, 1204-1206.
- Lebherz and Rutter (1973) J. Biol. Chem. 248:1650-1659.
- Lebherz et al (1984) J. Biol. Chem. 259 (2):1011-1017.
- Leyva et al. (1995) Plant Physiology 108(1):39-46.
- 20 London and Kline (1973) Bacteriol. Rev. 37:453-478.
- Lloyd et al. (1991). Mol. Gen. Genet. 225 (2):209-216.
- Luan et al. (1992). Plant Cell 4 (8):971-981.
- Lueberstedt et al. (1994) Plant Physiology 104 (3):997-1006.
- Maniatis et al. (1982) Molecular Cloning: A laboratory manual. Cold Spring Harbor Laboratory
- 25 Press, Cold Spring Harbor, N.Y.
- Matsuoka et al. (1993). Proc. Natl. Acad. Sci. U. S. A. 90(20):9586-9590.
- Mignery et al (1988) Gene 62:27-44.
- Muller et al (1990) Mol. Gen. Genet. 224:136-146.
- Oelmueller et al. (1992). Res. Photosynth., Proc. Int. Congr. Photosynth., 9th.
- 30 Volume 3: p219-24. Editor(s): Murata, Norio. Publisher: Kluwer, Dordrecht, Neth.

- Pedersen et al. (1982) Cell 29:1015-1026.
- Potrykus et al. (1985), Mol. Gen. Genet. 199:183-188.
- Rocha-Sosa et al. (1989) EMBO J. 8 (1):23-29.
- Rogers et al. (1987) Improved vectors for plant transformation: expression cassette vectors and
 5 new selectable markers. In Methods in Enzymology. Edited by R. Wu and L. Grossman.
 p253-277. San Diego: Academic Press.
- Rohde et al. (1990) J. Genet. & Breed. 44:311-315.
- Russell et al. (1993) Plant Cell Reports 13:24-27.
- Russell and Fromm (1997) Transgenic Research 6 (2):157-168.
- 10 Russel and Gibbs (1967) Biochim. Biophys. Acta 132, 145-154
- Salanoubat and Belliard (1987) Gene 60:47-56.
- Salanoubat and Belliard (1989) Gene 84:181-185.
- Samac et al. (1990) Plant Physiol. 93:907-914.
- Sambrook et al. (1989) Molecular cloning: A laboratory manual, 2nd ed., Cold Spring Harbor
 15 Laboratory Press, Cold Spring Harbor, N. Y.
- Schmidhauser and Helinski. (1985) J. Bacteriol. 164-155.
- Sonnewald et al. (1994) Plant Cell and Environment 17:649-658.
- Stalker et al. (1988) J. Biol. Chem. 263:6310-6314.
- Stark et al. (1992) Science 258: 287-292.
- 20 Stockhaus et al. (1989). EMBO Journal 8(9):2445-2451.
- Stribling and Perham (1973) Biochem. J. 131:833-841.
- Suzuki et al. (1994) Plant Mol. Biol. 25(3):507-516.
- Thillet et al. (1988) J. Biol. Chem. 263:12500-12508.
- Thompson et al. (1994) Nucl. Acids Res. 22:4673-4680.
- 25 Tierney et al. (1987) Planta 172:356-363.
- Truernit et al. (1995) Planta 196 (3):564-570.
- Tsutsumi et al. (1984) J. Biol. Chem. 259, 14572-14575.
- Umbeck et al. (1987) Biotechnology. 5, 263-266.
- Vain et al. (1993) Plant Cell Reports 12: 84-88.
- 30 Vasil et al. (1990) Bio/Technology 8:429-434.

Vasil et al. (1992) Bio/Technology 10:667-674.

Walters et al. (1992) Plant Molecular Biology 18: 189-200.

Witke and Goetz (1993) Journal of Bacteriology 175(22): 7495-7499

Wong et al. (1988) Gene 68: 193-203.

5 Yamamoto et al. (1994) Plant and Cell Physiology 35(5):773-778.

Zheng et al. (1993) Plant J. 4: 3357-3366.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Gerard Barry

NordineCheikh

Ganesh Kishore

(ii) TITLE OF INVENTION: Expression of Fructose 1,6 Bisphosphate Aldolase
in Transgenic Plants

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Arnold, White & Durkee

(B) STREET: P.O. Box 4433

(C) CITY: Houston

(D) STATE: Texas

(E) COUNTRY: United States of America

(F) ZIP: 77210-4433

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US Unknown

(B) FILING DATE: Concurrently Herewith

(C) CLASSIFICATION: Unknown

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US Prov. App. Serial No. 60/049,995

(B) FILING DATE: June 17, 1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Patricia A. Kammerer

(B) REGISTRATION NUMBER: 29,775

(C) REFERENCE/DOCKET NUMBER: MOBT086

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (713) 787-1400

(B) TELEFAX: (713) 787-1440

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1080 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGTCTAAGA TTTTGTATTT CGTAAAACCT GGCCTAATCA CTGGTGATGA CGTACAGAAA	60
GTTTTCCAGG TAGCAAAAGA AAACAACCTC GCACTGCCAG CAGTAAACTG CGTCGGTACT	120
GACTCCATCA ACGCCGTACT GGAAACCGCT GCTAAAGTTA AAGCGCCGGT TATCGTTCAG	180
TTCTCCAACG GTGGTGCTTC CTTTATCGCT GGTAAAGGCG TGAAATCTGA CGTTCCGCAG	240
GGTGCTGCTA TCCTGGGCGC GATCTCTGGT GCGCATCACG TTCACCAGAT GGCTGAACAT	300
TATGGTGTTT CGGTTATCCT GCACACTGAC CACTGCGCGA AGAAACTGCT GCCGTGGATC	360
GACGGTCTGT TGGACGCGGG TGAAAAACAC TTCGCAGCTA CCGGTAAGCC GCTGTTCTCT	420
TCTCACATGA TCGACCTGTC TGAAGAATCT CTGCAAGAGA ACATCGAAAT CTGCTCTAAA	480
TACCTGGAGC GCATGTCCAA AATCGGCATG ACTCTGGAAA TCGAACTGGG TTGCACCGGT	540
GGTGAAGAAG ACGGCGTGGA CAACAGCCAC ATGGACGCTT CTGCACTGTA CACCCAGCCG	600
GAAGACGTTG ATTACGCATA CACCGAACTG AGCAAAATCA GCCCGCGTTT CACCATCGCA	660
GCGTCCTTCG GTAACGTACA CGGTGTTTAC AAGCCGGGTA ACGTGGTTCT GACTCCGACC	720
ATCCTGCGTG ATTCTCAGGA ATATGTTTCC AAGAAACACA ACCTGCCGCA CAACAGCCTG	780
AACTTCGTAT TCCACGGTGG TTCCGGTTCT ACTGCTCAGG AAATCAAAGA CTCCGTAAGC	840
TACGGCGTAG TAAAAATGAA CATCGATACC GATACCCAAT GGGCAACCTG GGAAGGCGTT	900
CTGAACTACT ACAAAGCGAA CGAAGCTTAT CTGCAGGGTC AGCTGGGTAA CCCGAAAGGC	960
GAAGATCAGC CGAACAAGAA ATACTACGAT CCGCGCGTAT GGCTGCGTGC CGGTCAGACT	1020
TCGATGATCG CTCGTCTGGA GAAAGCATTC CAGGAACTGA ACGCGATCGA CGTTCTGTAA	1080

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 359 amino acids
 (B) TYPE: amino

(C) STRANDEDNESS:
 (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2

Met	Ser	Lys	Ile	Phe	Asp	Phe	Val	Lys	Pro	Gly	Val	Ile	Thr	Gly	5	10	15
Asp	Asp	Val	Gln	Lys	Val	Phe	Gln	Val	Ala	Lys	Glu	Asn	Asn	Phe	20	25	30
Ala	Leu	Pro	Ala	Val	Asn	Cys	Val	Gly	Thr	Asp	Ser	Ile	Asn	Ala	35	40	45
Val	Leu	Glu	Thr	Ala	Ala	Lys	Val	Lys	Ala	Pro	Val	Ile	Val	Gln	50	55	60
Phe	Ser	Asn	Gly	Gly	Ala	Ser	Phe	Ile	Ala	Gly	Lys	Gly	Val	Lys	65	70	75
Ser	Asp	Val	Pro	Gln	Gly	Ala	Ala	Ile	Leu	Gly	Ala	Ile	Ser	Gly	80	85	90
Ala	His	His	Val	His	Gln	Met	Ala	Glu	His	Tyr	Gly	Val	Pro	Val	95	100	105
Ile	Leu	His	Thr	Asp	His	Cys	Ala	Lys	Lys	Leu	Leu	Pro	Trp	Ile	110	115	120
Asp	Gly	Leu	Leu	Asp	Ala	Gly	Glu	Lys	His	Phe	Ala	Ala	Thr	Gly	125	120	135
Lys	Pro	Leu	Phe	Ser	Ser	His	Met	Ile	Asp	Leu	Ser	Glu	Glu	Ser	140	145	150
Leu	Gln	Glu	Asn	Ile	Glu	Ile	Cys	Ser	Lys	Tyr	Leu	Glu	Arg	Met	155	160	165
Ser	Lys	Ile	Gly	Met	Thr	Leu	Glu	Ile	Glu	Leu	Gly	Cys	Thr	Gly	170	175	180
Gly	Glu	Glu	Asp	Gly	Val	Asp	Asn	Ser	His	Met	Asp	Ala	Ser	Ala	185	190	195
Leu	Tyr	Thr	Gln	Pro	Glu	Asp	Val	Asp	Tyr	Ala	Tyr	Thr	Glu	Leu	200	205	210
Ser	Lys	Ile	Ser	Pro	Arg	Phe	Thr	Ile	Ala	Ala	Ser	Phe	Gly	Asn	215	220	225
Val	His	Gly	Val	Tyr	Lys	Pro	Gly	Asn	Val	Val	Leu	Thr	Pro	Thr	230	235	240

Ile	Leu	Arg	Asp	Ser	Gln	Glu	Tyr	Val	Ser	Lys	Lys	His	Asn	Leu	245	250	255
Pro	His	Asn	Ser	Leu	Asn	Phe	Val	Phe	His	Gly	Gly	Ser	Gly	Ser	260	265	270
Thr	Ala	Gln	Glu	Ile	Lys	Asp	Ser	Val	Ser	Tyr	Gly	Val	Val	Lys	275	280	285
Met	Asn	Ile	Asp	Thr	Asp	Thr	Gln	Trp	Ala	Thr	Trp	Glu	Gly	Val	290	295	300
Leu	Asn	Tyr	Tyr	Lys	Ala	Asn	Glu	Ala	Tyr	Leu	Gln	Gly	Gln	Leu	305	310	315
Gly	Asn	Pro	Lys	Gly	Glu	Asp	Gln	Pro	Asn	Lys	Lys	Tyr	Tyr	Asp	320	325	330
Pro	Arg	Val	Trp	Leu	Arg	Ala	Gly	Gln	Thr	Ser	Met	Ile	Ala	Arg	335	340	345
Leu	Glu	Lys	Ala	Phe	Gln	Glu	Leu	Asn	Ala	Ile	Asp	Val	Leu		350	355	

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGGGCCATGG CTAAGATTTT TGATTTCGTA

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CCCCGAGCTC TTACAGAACG TCGATCGCGT TCAG

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10847 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
1  CGATAAGCTT GATGTAATTG GAGGAAGATC AAAATTTTCA ATCCCCATTC
51  TTCGATTGCT TCAATTGAAG TTTCTCCGAT GCGCAAGTT AGCAGAATCT
101 GCAATGGTGT GCAGAACCCA TCTCTTATCT CCAATCTCTC GAAATCCAGT
151 CAACGCAAAT CTCCCTTATC GGTCTTCTCTG AAGACGCAGC AGCATCCACG
201 AGCTTATCCG ATTTTCGTCGT CGTGGGGATT GAAGAAGAGT GGGATGACGT
251 TAATTGGCTC TGAGCTTCGT CCTCTTAAGG TCATGTCTTC TGTTCACG
301 GCGTGCATGC TTCACGGTGC AAGCAGCCGT CCAGCAACTG CTCGTAAGTC
351 CTCTGGTCTT TCTGGAACCG TCCGTATTCC AGGTGACAAG TCTATCTCCC
401 ACAGGTCCTT CATGTTTGGG GGTCTCGCTA GCGGTGAAAC TCGTATCACC
451 GGTCTTTTGG AAGGTGAAGA TGTTATCAAC ACTGGTAAGG CTATGCAAGC
501 TATGGGTGCC AGAATCCGTA AGGAAGGTGA TACTTGGATC ATTGATGGTG
551 TTGGTAACGG TGGACTCCTT GCTCCTGAGG CTCCTCTCGA TTTCCGTAAC
601 GCTGCAACTG GTTGCCGTTT GACTATGGGT CTTGTTGGTG TTTACGATTT
651 CGATAGCACT TTCATTGGTG ACGCTTCTCT CACTAAGCGT CCAATGGGTC
701 GTGTGTTGAA CCCACTTCGC GAAATGGGTG TGCAGGTGAA GTCTGAAGAC
751 GGTGATCGTC TTCCAGTTAC CTTGCGTGGA CCAAAGACTC CAACGCCAAT
801 CACCTACAGG GTACCTATGG CTTCCGCTCA AGTGAAGTCC GCTGTTCTGC
851 TTGCTGGTCT CAACACCCCA GGTATCACCA CTGTTATCGA GCCAATCATG
901 ACTCGTGACC AACTGAAAAA GATGCTTCAA GGTTTTGGTG CTAACCTTAC
951 CGTTGAGACT GATGCTGACG GTGTGCGTAC CATCCGTCTT GAAGGTCGTG
1001 GTAAGCTCAC CGGTCAAGTG ATTGATGTTT CAGGTGATCC ATCCTCTACT
1051 GCTTTCCCAT TGGTTGCTGC CTTGCTTGTT CCAGGTTCCG ACGTCACCAT
1101 CCTTAACGTT TTGATGAACC CAACCCGTAC TGGTCTCATC TTGACTCTGC
1151 AGGAAATGGG TGCCGACATC GAAGTGATCA ACCCACGTCT TGCTGGTGGA
1201 GAAGACGTGG CTGACTTGCG TGTTCTGTTCT TCTACTTTGA AGGGTGTTAC
1251 TGTTCCAGAA GACCGTGCTC CTTCTATGAT CGACGAGTAT CCAATTCTCG
1301 CTGTTGCAGC TGCATTGCTT GAAGGTGCTA CCGTTATGAA CGGTTTGGAA
1351 GAAGTCCGTG TTAAGGAAAG CGACCGTCTT TCTGCTGTCG CAAACGGTCT
1401 CAAGCTCAAC GGTGTTGATT GCGATGAAGG TGAGACTTCT CTCGTCGTGC
1451 GTGGTCTGCC TGACGGTAAG GGTCTCGGTA ACGCTTCTGG AGCAGCTGTC
1501 GCTACCCACC TCGATCACCG TATCGCTATG AGCTTCCTCG TTATGGGTCT
1551 CGTTTCTGAA AACCCTGTTA CTGTTGATGA TGCTACTATG ATCGCTACTA
1601 GCTTCCAGAA GTTCATGGAT TTGATGGCTG GTCTTGGAGC TAAGATCGAA
1651 CTCTCCGACA CTAAGGCTGC TTGATGAGCT CAAGAATTCG AGCTCGGTAC
1701 CGGATCCAGC TTTTCGTTCT ATCATCGGTT TCGACAACGT TCGTCAAGTT
1751 CAATGCATCA GTTTCATTGC GCACACACCA GAATCCTACT GAGTTCGAGT
1801 ATTATGGCAT TGGGAAAAC TTTTTTCTTG TACCATTGTG TGTGCTTGTA
1851 ATTTACTGTG TTTTTTATTC GGTTCGCTAT ATCGAACTGT GAAATGGAAA
1901 TGGATGGAGA AGAGTTAATG AATGATATGG TCCTTTTGTG CATTCTCAAA
1951 TTAATATTAT TTGTTTTTTC TCTTATTTGT TGTGTGTTGA ATTTGAAATT
2001 ATAAGAGATA TGCAAACATT TTGTTTTGAG TAAAAATGTG TCAAATCGTG
2051 GCCTCTAATG ACCGAAGTTA ATATGAGGAG TAAACACTT GTAGTTGTAC
2101 CATTATGCTT ATTCACTAGG CAACAAATAT ATTTTCAGAC CTAGAAAAGC
2151 TGCAAATGTT ACTGAATACA AGTATGTCCT CTTGTGTTTT AGACATTTAT
2201 GAACTTTCCT TTATGTAATT TTCCAGAATC CTTGTCAGAT TCTAATCATT
2251 GCTTTATAAT TATAGTTATA CTCATGGATT TGAGTTGAG TATGAAAATA
2301 TTTTTTAATG CATTTTATGA CTTGCCAATT GATTGACAAC ATGCATCAAT
2351 CGACCTGCAG CCACTCGAAG CGGCCGCGTT CAAGCTTGAG CTCAGGATTT
2401 AGCAGCATT CAGATTGGGT TCAATCAACA AGGTACGAGC CATATCACTT
```

2451 TATTCAAATT GGTATCGCCA AAACCAAGAA GGAAC TCCCA TCCTCAAAGG
2501 TTTGTAAGGA AGAATTCTCA GTCCAAAGCC TCAACAAGGT CAGGGTACAG
2551 AGTCTCCAAA CCATTAGCCA AAAGCTACAG GAGATCAATG AAGAATCTTC
2601 AATCAAAGTA AACTACTGTT CCAGCACATG CATCATGGTC AGTAAGTTTC
2651 AGAAAAAGAC ATCCACCGAA GACTTAAAGT TAGTGGGCAT CTTTGAAAGT
2701 AATCTTGTC AATCGAGCA GCTGGCTTGT GGGGACCAGA CAAAAAAGGA
2751 ATGGTGCAGA ATTGTTAGGC GCACCTACCA AAAGCATCTT TGCCTTTATT
2801 GCAAAGATAA AGCAGATTCC TCTAGTACAA GTGGGGAACA AAATAACGTG
2851 GAAAAGAGCT GTCCTGACAG CCCACTCACT AATGCGTATG ACGAACGCAG
2901 TGACGACCAC AAAAGAATTC CCTCTATATA AGAAGGCATT CATTCCCATT
2951 TGAAGGATCA TCAGATACTG AACCAATCCT TCTAGAACTT CTCCACAATG
3001 GCTTCCTCTA TGCTCTCTTC CGCTACTATG GTTGCCTCTC CGGCTCAGGC
3051 CACTATGGTC GTCCTTTTCA ACGGACTTAA GTCCTCCGCT GCCTTCCCAG
3101 CCACCCGCAA GGCTAACAA GACATTACTT CCATCACAAG CAACGGCGGA
3151 AGAGTTAACT GCATGCAGGT GTGGCCTCCG ATTGGAAAGA AGAAGTTTGA
3201 GACTCTCTCT TACCTTCCTG ACCTTACCGA TTCCGGTGGT CGCGTCAACT
3251 GCATGCAGGC CATGGCTAAG ATTTTGTGATT TCGTAAAACC TGGCGTAATC
3301 ACTGGTGATG ACGTACAGAA AGTTTTCCAG GTAGCAAAAG AAAACAACCTT
3351 CGCACTGCCA GCAGTAAACT GCGTCGGTAC TGAATCCATC AACGCCGTAC
3401 TGGAAACCGC TGCTAAAGTT AAAGCGCCGG TTATCGTTCA GTTCTCCAAC
3451 GGTGGTGCTT CCTTTATCGC TGGTAAAGGC GTGAAATCTG ACGTTCCGCA
3501 GGGTGCTGCT ATCCTGGGCG CGATCTCTGG TGCGCATCAC GTTACCAGGA
3551 TGGCTGAACA TTATGGTGTT CCGGTTATCC TGCACACTGA CCACTGCGCG
3601 AAGAACTGCT TGCCGTGGAT CGACGGTCTG TTGGACGCGG GTGAAAAACA
3651 CTTGCGAGCT ACCGGTAAGC CGCTGTTCTC TTCTCACATG ATCGACCTGT
3701 CTGAAGAATC TCTGCAAGAG AACATCGAAA TCTGCTCTAA ATACCTGGAG
3751 CGCATGTCCA AAATCGGCAT GACTCTGGAA ATCGAACTGG GTTGCACCGG
3801 TGGTGAAGAA GACGGCGTGG ACAACAGCCA CATGGACGCT TCTGCACTGT
3851 ACACCCAGCC GGAAGACGTT GATTACGCAT ACACCGAACT GAGCAAAATC
3901 AGCCCGCGTT TCACCATCGC AGCGTCCTTC GGTAACGTAC ACGGTGTTTA
3951 CAAGCCGGGT AACGTGGTTC TGAATCCGAC CATCCTGCGT GATTCTCAGG
4001 AATATGTTTC CAAGAAACAC AACCTGCCGC ACAACAGCCT GAACTTCGTA
4051 TTCCACGGTG GTTCCGGTTC TACTGCTCAG GAAATCAAAG ACTCCGTAAG
4101 CTACGGCGTA GTAAAAATGA ACATCGATAC CGATACCCAA TGGGCAACCT
4151 GGGAAAGCGT TCTGAACTAC TACAAAGCGA ACGAAGCTTA TCTGCAGGGT
4201 CAGCTGGGTA ACCCGAAAGG CGAAGATCAG CCGAACAAGA AATACTACGA
4251 TCCGCGCGTA TGGCTGCGTG CCGGTCAGAC TTCGATGATC GCTCGTCTGG
4301 AGAAAGCATT CCAGGAAC TG AACGCGATCG ACGTTCTGTA AGAGCTCGGT
4351 ACCGGATCCA ATTCCCGATC GTTCAAACAT TTGGCAATAA AGTTTCTTAA
4401 GATTGAATCC TGTTGCCGGT CTTGCGATGA TTATCATATA ATTTCTGTTG
4451 AATTACGTTA AGCATGTAAT AATTAACATG TAATGCATGA CGTTATTTAT
4501 GAGATGGGTT TTTATGATTA GAGTCCCGCA ATTATACATT TAATACGCGA
4551 TAGAAAACAA AATATAGCGC GCAAAC TAGG ATAAATTATC GCGCGCGGTG
4601 TCATCTATGT TACTAGATCG GGGATCGATC CCCGGGCGGC CGCCACTCGA
4651 GTGGTGGCCG CATCGATCGT GAAGTTTCTC ATCTAAGCCC CCATTTGGAC
4701 GTGAATGTAG ACACGTCGAA ATAAAGATTT CCGAATTAGA ATAATTTGTT
4751 TATTGCTTTC GCCTATAAAT ACGACGGATC GTAATTTGTC GTTTTATCAA
4801 AATGTACTTT CATTTTATAA TAACGCTGCG GACATCTACA TTTTGAATT
4851 GAAAAAAAT TGGTAATTAC TCTTTCTTTT TCTCCATATT GACCATCATA
4901 CTCATTGCTG ATCCATGTAG ATTTCCCGGA CATGAAGCCA TTTACAATTG
4951 AATATATCCT GCCGCCGCTG CCGCTTTGCA CCCGGTGGAG CTTGCATGTT
5001 GGTTCCTACG CAGAACTGAG CCGGTTAGGC AGATAATTTT CATTGAGAAC
5051 TGAGCCATGT GCACCTTCCC CCAACACGG TGAGCGACGG GGCAACGGAG
5101 TGATCCACAT GGGACTTTTc CTAGCTTGGC TGCCATTTTT GGGGTGAGGC

5151	CGTTCGCGCG	GGGCGCCAGC	TGGGGGGGATG	GGAGGCCCGC	GTTACCGGGA
5201	GGGTTCGAGA	AGGGGGGGCA	CCCCCCTTCG	GCGTGCGCGG	TCACGCGCCA
5251	GGGCGCAGCC	CTGGTTAAAA	ACAAGGTTTA	TAAATATTGG	TTTAAAAGCA
5301	GGTTAAAAAG	CAGGTTAGCG	GTGGCCGAAA	AACGGGCGGA	AACCCTTGCA
5351	AATGCTGGAT	TTTCTGCCTG	TGGACAGCCC	CTCAAATGTC	AATAGGTGCG
5401	CCCCTCATCT	GTCATCACTC	TGCCCCCTCA	GTGTCAAGGA	TCGCGCCCCCT
5451	CATCTGTCAG	TAGTCGCGCC	CCTCAAGTGT	CAATACCGCA	GGGCACTTAT
5501	CCCCAGGCTT	GTCCACATCA	TCTGTGGGAA	ACTCGCGTAA	AATCAGGCGT
5551	TTTCGCCGAT	TTGCGAGGCT	GGCCAGCTCC	ACGTGCGCCG	CCGAAATCGA
5601	GCCTGCCCTT	CATCTGTCAA	CGCCGCGCCG	GGTGAGTCGG	CCCCTCAAGT
5651	GTCAACGTCC	GCCCCCTCAT	TGTCAGTGAG	GGCCAAGTTT	TCCGCGTGGT
5701	ATCCACAACG	CCGGCGGCCG	GCCGCGGTGT	CTCGCACACG	GCTTCGACGG
5751	CGTTTCTGGC	GCGTTTGCAG	GGCCATAGAC	GGCCGCCAGC	CCAGCGGCCGA
5801	GGGCAACCAG	CCCGGTGAGC	GTCGGAAAGG	GTCGATCGAC	CGATGCCCTT
5851	GAGAGCCTTC	AACCCAGTCA	GCTCCTTCCG	GTGGGCGCGG	GGCATGACTA
5901	TCGTGCGCCG	ACTTATGACT	GTCTTCTTTA	TCATGCAACT	CGTAGGACAG
5951	GTGCCGGCAG	CGCTCTGGGT	CATTTTCGGC	GAGGACCGCT	TTCGCTGGAG
6001	CGCGACGATG	ATCGGCCTGT	CGCTTGCGGT	ATTCGGAATC	TTGCACGCCC
6051	TCGCTCAAGC	CTTCGTCACT	GGTCCCGCCA	CCAAACGTTT	CGGCGAGAAG
6101	CAGGCCATTA	TCGCCGGCAT	GGCGGCCGAC	GCGCTGGGCT	ACGTCTTGCT
6151	GGCGTTTCGG	ACGCGAGGCT	GGATGGCCTT	CCCCATTATG	ATTCTTCTCG
6201	CTTCCGGCGG	CATCGGGATG	CCCGCGTTGC	AGGCCATGCT	GTCCAGGCAG
6251	GATAGTACG	ACCATCAGGG	ACAGCTTCAA	GGATCGCTCG	CGGCTCTTAC
6301	CAGCCTAACT	TCGATCACTG	GACCGCTGAT	CGTCACGGCG	ATTTATGCCG
6351	CCTCGGCGAG	CACATGGAAC	GGGTTGGCAT	GGATTGTAGG	CGCCGCCCTA
6401	TACCTTGTC	GCCTCCCCGC	GTTGCGTCGC	GGTGATGGA	GCCGGGCCAC
6451	CTCGACCTGA	ATGGAAGCCG	GCGGCACCTC	GCTAACGGAT	TCACCACTCC
6501	AAGAATTGGA	GCCAATCAAT	TCTTGCGGAG	AACTGTGAAT	GCGCAAACCA
6551	ACCCTTGGCA	GAACATATCC	ATCGCGTCCG	CCATCTCCAG	CAGCCGCACG
6601	CGGCGCATCT	CGGGCAGCGT	TGGGTCTTGG	CCACGGGTGC	GCATGATCGT
6651	GCTCCTGTCT	TTGAGGACCC	GGCTAGGCTG	GCGGGGTGTC	CTTACTGGTT
6701	AGCAGAATGA	ATCACCGATA	CGCGAGCGAA	CGTGAAGCGA	CTGCTGCTGC
6751	AAAACGTCTG	CGACCTGAGC	AACAACATGA	ATGGTCTTCG	GTTTCCGTGT
6801	TTTCGTAAAG	CTGGAAACGC	GGAAGTCAGC	GCCCTGCACC	ATTATGTTCC
6851	GGATCTGCAT	CGCAGGATGC	TGCTGGCTAC	CCTGTGGAAC	ACCTACATCT
6901	GTATTAACGA	AGCGCTGGCA	TTGACCCTGA	GTGATTTTTC	TCTGGTCCCC
6951	CCGCATCCAT	ACCGCCAGTT	GTTTACCCTC	ACAACGTTCC	AGTAACCGGG
7001	CATGTTTCAT	ATCAGTAACC	CGTATCGTGA	GCATCCTCTC	TCGTTTCATC
7051	GGTATCATTA	CCCCCATGAA	CAGAAATTCC	CCCTTACACG	GAGGCATCAA
7101	GTGACCAAAC	AGGAAAAAAC	CGCCCTTAAC	ATGGCCCGCT	TTATCAGAAG
7151	CCAGACATTA	ACGCTTCTGG	AGAAACTCAA	CGAGCTGGAC	GCGGATGAAC
7201	AGGCAGACAT	CTGTGAATCG	CTTCACGACC	ACGCTGATGA	GCTTTACCGC
7251	AGCTGCCTCG	CGCGTTTCGG	TGATGACGGT	GAAAACCTCT	GACACATGCA
7301	GCTCCCGGAG	ACGGTCACAG	CTTGTCTGTA	AGCGGATGCC	GGGAGCAGAC
7351	AAGCCCGTCA	GGGCGCGTCA	GCGGGTGTTG	GCGGGTGTCG	GGGCGCAGCC
7401	ATGACCCAGT	CACGTAGCGA	TAGCGGAGTG	TATACTGGCT	TAAGTATGCG
7451	GCATCAGAGC	AGATTGTACT	GAGAGTGCAC	CATATGCGGT	GTGAAATACC
7501	GCACAGATGC	GTAAGGAGAA	AATACCGCAT	CAGGCGCTCT	TCCGCTTCCT
7551	CGCTCACTGA	CTCGCTGCGC	TCGGTTCGTT	GGCTGCGGCG	AGCGGTATCA
7601	GCTCACTCAA	AGGCGGTAAT	ACGGTTATCC	ACAGAATCAG	GGGATAACGC
7651	AGGAAAGAAC	ATGTGAGCAA	AAGGCCAGCA	AAAGGCCAGG	AACCGTAAAA
7701	AGGCCGCGTT	GCTGGCGTTT	TTCCATAGGC	TCCGCCCCCC	TGACGAGCAT
7751	CACAAAAATC	GACGCTCAAG	TCAGAGGTGG	CGAAACCCGA	CAGGACTATA
7801	AAGATACCAG	GCGTTTCCCC	CTGGAAGCTC	CCTCGTGCGC	TCTCCTGTTC

7851	CGACCCTGCC	GCTTACCGGA	TACCTGTCCG	CCTTTCTCCC	TTCGGGAAGC
7901	GTGGCGCTTT	CTCATAGCTC	ACGCTGTAGG	TATCTCAGTT	CGGTGTAGGT
7951	CGTTTCGCTCC	AAGCTGGGCT	GTGTGCACGA	ACCCCCCGTT	CAGCCCCGACC
8001	GCTGCGCCTT	ATCCGGTAAC	TATCGTCTTG	AGTCCAACCC	GGTAAGACAC
8051	GACTTATCGC	CACTGGCAGC	AGCCACTGGT	AACAGGATTA	GCAGAGCGAG
8101	GTATGTAGGC	GGTGCTACAG	AGTTCTTGAA	GTGGTGGCCT	AACTACGGCT
8151	ACACTAGAAG	GACAGTATTT	GGTATCTGCG	CTCTGCTGAA	GCCAGTTACC
8201	TTCGGAAGAA	GAGTTGGTAG	CTCTTGATCC	GGCAAACAAA	CCACCGCTGG
8251	TAGCGGTGGT	TTTTTTGTTT	GCAAGCAGCA	GATTACGCGC	AGAAAAAAG
8301	GATCTCAAGA	AGATCCTTTG	ATCTTTTCTA	CGGGGTCTGA	CGCTCAGTGG
8351	AACGAAAACT	CACGTTAAGG	GATTTTGGTC	ATGAGATTAT	CAAAAAGGAT
8401	CTTCACCTAG	ATCCTTTTAA	ATTAAAAATG	AAGTTTAAAA	TCAATCTAAA
8451	GTATATATGA	GTAAACTTGG	TCTGACAGTT	ACCAATGCTT	AATCAGTGAG
8501	GCACCTATCT	CAGCGATCTG	TCTATTTCTG	TCATCCATAG	TTGCCCTGACT
8551	CCCCGTCGTG	TAGATAACTA	CGATACGGGA	GGGCTTACCA	TCTGGCCCCA
8601	GTGCTGCAAT	GATACCGCGA	GACCCACGCT	CACCGGCTCC	AGATTTATCA
8651	GCAATAAAACC	AGCCAGCCGG	AAGGGCCGAG	CGCAGAAGTG	GTCTTGCAAC
8701	TTTATCCGCC	TCCATCCAGT	CTATTAATTG	TTGCCGGGAA	GCTAGAGTAA
8751	GTAGTTTCGCC	AGTTAATAGT	TTGCGCAACG	TTGTTGCCAT	TGCTGCAGGT
8801	CGGGAGCACA	GGATGACGCC	TAACAATTCA	TTCAAGCCGA	CACCGCTTCG
8851	CGGCGCGGCT	TAATTCAAGG	GTTAAACATC	ATGAGGGAAG	CGGTGATCGC
8901	CGAAGTATCG	ACTCAACTAT	CAGAGGTAGT	TGGCGTCATC	GAGCGCCATC
8951	TCGAACCGAC	GTTGCTGGCC	GTACATTTGT	ACGGCTCCGC	AGTGGATGGC
9001	GGCCTGAAGC	CACACAGTGA	TATTGATTTG	CTGGTTACGG	TGACCGTAAG
9051	GCTTGATGAA	ACAACGCGGC	GAGCTTTGAT	CAACGACCTT	TTGGAAACTT
9101	CGGCTTCCCC	TGGAGAGAGC	GAGATTCTCC	GCGCTGTAGA	AGTCACCATT
9151	GTTGTGCACG	ACGACATCAT	TCCGTGGCGT	TATCCAGCTA	AGCGCGAACT
9201	GCAATTTGGA	GAATGGCAGC	GCAATGACAT	TCTTGACAGT	ATCTTCGAGC
9251	CAGCCACGAT	CGACATTGAT	CTGGCTATCT	TGCTGACAAA	AGCAAGAGAA
9301	CATAGCGTTG	CCTTGGTAGG	TCCAGCGGCG	GAGGAACTCT	TTGATCCGGT
9351	TCCTGAACAG	GATCTATTTG	AGGCGCTAAA	TGAAACCTTA	ACGCTATGGA
9401	ACTCGCCGCC	CGACTGGGCT	GGCGATGAGC	GAAATGTAGT	GCTTACGTTG
9451	TCCCGCATTT	GGTACAGCGC	AGTAACCGGC	AAAAATCGCGC	CGAAGGATGT
9501	CGCTGCCGAC	TGGGCAATGG	AGCGCCTGCC	GGCCAGTAT	CAGCCCGTCA
9551	TACTTGAAAGC	TAGGCAGGCT	TATCTTGGAC	AAGAAGATCG	CTTGGCCTCG
9601	CGCGCAGATC	AGTTGGAAGA	ATTTGTTCAC	TACGTGAAAG	GCGAGATCAC
9651	CAAGGTAGTC	GGCAAATAAT	GTCTAACAAT	TCGTTCAAGC	CGACGCCGCT
9701	TCGCGGCGCG	GCTTAACTCA	AGCGTTAGAT	GCTGCAGGCA	TCGTGGTGTG
9751	ACGCTCGTCG	TTTGGTATGG	CTTCATTCAG	CTCCGGTTCC	CAACGATCAA
9801	GGCGAGTTAC	ATGATCCCCC	ATGTTGTGCA	AAAAAGCGGT	TAGCTCCTTC
9851	GGTCTCCGA	TCGAGGATTT	TTGCGGCTG	CGCTACGTCC	GCKACCGCGT
9901	TGAGGGATCA	AGCCACAGCA	GCCCACTCGA	CCTCTAGCCG	ACCCAGACGA
9951	GCCAAGGGAT	CTTTTTGGAA	TGCTGCTCCG	TCGTCAGGCT	TTCCGACGTT
10001	TGGGTGGTTG	AACAGAAGTC	ATTATCGTAC	GGAATGCCAA	GCACTCCCGA
10051	GGGGAACCCCT	GTGGTTGGCA	TGCACATACA	AATGGACGAA	CGGATAAACC
10101	TTTTACGCCC	CTTTTAAATA	TCCGTTATTC	TAATAAACGC	TCTTTTCTCT
10151	TAGGTTTACC	CGCCAATATA	TCCTGTCAAA	CACTGATAGT	TTAAACTGAA
10201	GGCGGGAAAC	GACAATCTGA	TCCCCATCAA	GCTTGAGCTC	AGGATTTAGC
10251	AGCATTCAG	ATTGGGTTCA	ATCAACAAGG	TACGAGCCAT	ATCACTTTAT
10301	TCAAAATTGGT	ATCGCCAAAA	CCAAGAAGGA	ACTCCCATCC	TCAAAGGTTT
10351	GTAAGGAAGA	ATTCTCAGTC	CAAAGCCTCA	ACAAGGTCAG	GGTACAGAGT
10401	CTCCAAACCA	TTAGCCAAAA	GCTACAGGAG	ATCAATGAAG	AATCTTCAAT
10451	CAAAGTAAAC	TACTGTTCCA	GCACATGCAT	CATGGTCAGT	AAGTTTCAGA
10501	AAAAGACATC	CACCGAAGAC	TTAAAGTTAG	TGGGCATCTT	TGAAAGTAAT

```

10551 CTTGTCAACA TCGAGCAGCT GGCTTGTGGG GACCAGACAA AAAAGGAATG
10601 GTGCAGAATT GTTAGGCGCA CCTACCAAAA GCATCTTTGC CTTTATTGCA
10651 AAGATAAAGC AGATTCTCTCT AGTACAAGTG GGAACAAAA TAACGTGGAA
10701 AAGAGCTGTC CTGACAGCCC ACTCCTAAT GCGTATGACG AACGCAGTGA
10751 CGACCACAAA AGAATTCCCT CTATATAAGA AGGCATTCAT TCCCATTGTA
10801 AGGATCATCA GATACTGAAC CAATCCTTCT AGAAGATCTA AGCTTAT

```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10901 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```

1 CGATAAGCTT GATGTAATTG GAGGAAGATC AAAATTTTCA ATCCCCATTC
51 TTCGATTGCT TCAATTGAAG TTTCTCCGAT GCGCAAGTT AGCAGAATCT
101 GCAATGGTGT GCAGAACCCA TCTCTTATCT CCAATCTCTC GAAATCCAGT
151 CAACGCAAAAT CTCCCTTATC GGTTTCTCTG AAGACGCAGC AGCATCCACG
201 AGCTTATCCG ATTTCGTCGT CGTGGGGATT GAAGAAGAGT GGGATGACGT
251 TAATTGGCTC TGAGCTTCGT CCTCTTAAGG TCATGTCTTC TGTTCACG
301 GCGTGCAATG TTCACGGTGC AAGCAGCCGT CCAGCAACTG CTCGTAAGTC
351 CTCTGGTCTT TCTGGAACCG TCCGTATTCC AGGTGACAAG TCTATCTCCC
401 ACAGGTCCTT CATGTTTGGA GGTCTCGCTA GCGGTGAAAC TCGTATCACC
451 GGTCTTTTGG AAGGTGAAGA TGTTATCAAC ACTGGTAAGG CTATGCAAGC
501 TATGGGTGCC AGAATCCGTA AGGAAGGTGA TACTTGATC ATTGATGGTG
551 TTGGTAACGG TGGACTCCTT GCTCCTGAGG CTCCTCTCGA TTTCGGTAAC
601 GCTGCAACTG GTTGCCGTTT GACTATGGGT CTTGTTGGTG TTTACGATTT
651 CGATAGCACT TTCATTGGTG ACGCTTCTCT CACTAAGCGT CCAATGGGTC
701 GTGTGTTGAA CCCACTTCGC GAAATGGGTG TGCAAGTGAA GTCTGAAGAC
751 GGTGATCGTC TTCCAGTTAC CTTGCGTGGA CCAAAGACTC CAACGCCAAT
801 CACCTACAGG GTACCTATGG CTTCCGCTCA AGTGAAGTCC GCTGTTCTGC
851 TTGCTGGTCT CAACACCCCA GGTATCACCA CTGTTATCGA GCCAATCATG
901 ACTCGTGACC AACTGAAAA GATGCTTCAA GGTTTGGTG CTAACCTTAC
951 CGTTGAGACT GATGCTGACG GTGTGCGTAC CATCCGTCTT GAAGGTCGTG
1001 GTAAGCTCAC CGGTCAAGTG ATTGATGTTT CAGGTGATCC ATCCTCTACT
1051 GCTTTCCCAT TGGTTGCTGC CTTGCTTGTT CCAGGTTCCG ACGTCACCAT
1101 CCTTAACGTT TTGATGAACC CAACCCGTAC TGGTCTCATC TTGACTCTGC
1151 AGGAAATGGG TGCCGACATC GAAGTGATCA ACCCAGTCT TGCTGGTGGA
1201 GAAGACGTGG CTGACTTGCG TGTTCTGTTT TCTACTTTGA AGGGTGTTAC
1251 TGTTCAGAA GACCGTGCTC CTTCTATGAT CGACGAGTAT CCAATTCTCG
1301 CTGTTGCAGC TGCAATTCGCT GAAGGTGCTA CCGTTATGAA CGGTTTGGAA
1351 GAACTCCGTG TTAAGGAAAG CGACCGTCTT TCTGCTGTCG CAAACGGTCT
1401 CAAGCTCAAC GGTGTTGATT GCGATGAAGG TGAGACTTCT CTCGTCGTGC
1451 GTGGTCTGCC TGACGGTAAG GGTCTCGGTA ACGCTTCTGG AGCAGCTGTC
1501 GCTACCCACC TCGATCACCG TATCGCTATG AGCTTCCTCG TTATGGGTCT
1551 CGTTTCTGAA AACCCTGTTA CTGTTGATGA TGCTACTATG ATCGCTACTA
1601 GCTTCCAGAA GTTCATGGAT TTGATGGCTG GTCTTGAGC TAAGATCGAA
1651 CTCTCCGACA CTAAGGCTGC TTGATGAGCT CAAGAATTCG AGCTCGGTAC
1701 CGGATCCAGC TTTCGTTCTG ATCATCGGTT TCGACAACGT TCGTCAAGTT
1751 CAATGCATCA GTTTCATTGC GCACACACCA GAATCCTACT GAGTTCGAGT
1801 ATTATGGCAT TGGGAAAAC GTTTTCTTGG TACCATTGTG TGTGCTTGTA

```

1851 ATTTACTGTG TTTTTTATTC GGTTCCTCGCT ATCGAACTGT GAAATGGAAA
1901 TGGATGGAGA AGAGTTAATG AATGATATGG TCCTTTTGTT CATTCTCAAA
1951 TTAATATTAT TTGTTTTTTC TCTTATTTGT TGTGTGTTGA ATTTGAAATT
2001 ATAAGAGATA TGCAAACATT TTGTTTTGAG TAAAAATGTG TCAAATCGTG
2051 GCCTCTAATG ACCGAAGTTA ATATGAGGAG TAAACACTT GTAGTTGTAC
2101 CATTATGCTT ATTCACTAGG CAACAAATAT ATTTTCAGAC CTAGAAAAGC
2151 TGCAAATGTT ACTGAATACA AGTATGTCCT CTTGTGTTTT AGACATTTAT
2201 GAACTTTCCT TTATGTAATT TTCCAGAATC CTTGTCAGAT TCTAATCATT
2251 GCTTTATAAT TATAGTTATA CTCATGGATT TGTAGTTGAG TATGAAAATA
2301 TTTTTTAATG CATTTTATGA CTTGCCAATT GATTGACAAC ATGCATCAAT
2351 CGACCTGCAG CCACTCGAAG CGGCCGCGTT CAAGCTTGAG CTCAGGATTT
2401 AGCAGCATTC CAGATTGGGT TCAATCAACA AGGTACGAGC CATATCACTT
2451 TATTCAAATT GGTATCGCCA AAACCAAGAA GGAACCTCCA TCCTCAAAGG
2501 TTTGTAAGGA AGAATTCTCA GTCCAAAGCC TCAACAAGGT CAGGGTACAG
2551 AGTCTCCAAA CCATTAGCCA AAAGCTACAG GAGATCAATG AAGAATCTTC
2601 AATCAAAGTA AACTACTGTT CCAGCACATG CATCATGGTC AGTAAGTTTC
2651 AGAAAAAGAG ATCCACCGAA GACTTAAAGT TAGTGGGCAT CTTTGAAAGT
2701 AATCTTGTCA ACATCGAGCA GCTGGCTTGT GGGGACCAGA CAAAAAGGA
2751 ATGGTGCAGA ATTGTTAGGC GCACCTACCA AAAGCATCTT TGCCTTTATT
2801 GCAAAGATAA AGCAGATTCC TCTAGTACAA GTGGGGAACA AAATAACGTG
2851 GAAAAGAGCT GTCCTGACAG CCCACTCACT AATGCGTATG ACGAACGCAG
2901 TGACGACCAC AAAAGAATTC CCTCTATATA AGAAGGCATT CATTCCCATT
2951 TGAAGGATCA TCAGATACTG AACCAATCCT TCTAGAAGAT CTAAGCTTAT
3001 CGATAAGCTT GATGTAATTG GAGGAAGATC AAAATTTTCA ATCCCCATTC
3051 TTCGATTGCT TCAATTGAAG TTTCTCCGAT GCGCAAGTT AGCAGAATCT
3101 GCAATGGTGT GCAGAACCCA TCTCTTATCT CCAATCTCTC GAAATCCAGT
3151 CAACGCAAAT CTCCCTTATC GGTTCCTCTG AAGACGCAGC AGCATCCACG
3201 AGCTTATCCG ATTTCTGTCG CTGGGGGATT GAAGAAGAGT GGGATGACGT
3251 TAATTGGCTC TGAGCTTCGT CCTCTTAAGG TCATGTCTTC TGTTCACG
3301 GCGTGCATGC AGGCcatggc TAAGATTTTT GATTTCGTAA AACCTGGCGT
3351 AATCACTGGT GATGACGTAC AGAAAGTTTT CCAGGTAGCA AAAGAAAACA
3401 ACTTCGCACT GCCAGCAGTA AACTGCGTCG GTACTGACTC CATCAACGCC
3451 GTACTGGAAA CCGCTGCTAA AGTTAAAGCG CCGGTTATCG TTCAGTTCTC
3501 CAACGGTGGT GCTTCCTTTA TCGCTGGTAA AGGCGTGAAA TCTGACGTTT
3551 CGCAGGGTGC TGCTATCCTG GCGCGATCT CTGGTGCGCA TCACGTTTAC
3601 CAGATGGCTG AACATTATGG TGTTCCGGTT ATCCTGCACA CTGACCACTG
3651 CGCGAAGAAA CTGCTGCCGT GGATCGACGG TCTGTTGGAC GCGGGTGAAA
3701 AACACTTCGC AGCTACCGGT AAGCCGCTGT TCTCTTCTCA CATGATCGAC
3751 CTGTCTGAAG AATCTCTGCA AGAGAACATC GAAATCTGCT CTAAATACCT
3801 GGAGCGCATG TCCAAAATCG GCATGACTCT GGAAATCGAA CTGGGTTGCA
3851 CCGGTGGTGA AGAAGACGGC GTGGACAACA GCCACATGGA CGCTTCTGCA
3901 CTGTACACCC AGCCGGAAGA CGTTGATTAC GCATACACCG AACTGAGCAA
3951 AATCAGCCCG CGTTTCACCA TCGCAGCGTC CTTGCGTAAC GTACACGGTG
4001 TTTACAAGCC GGGTAACGTG GTTCTGACTC CGACCATCCT GCGTGATTCT
4051 CAGGAATATG TTTCCAAGAA ACACAACCTG CCGCACAACA GCCTGAACTT
4101 CGTATTCCAC GGTGGTTCCG GTTCTACTGC TCAGGAAATC AAAGACTCCG
4151 TAAGCTACGG CGTAGTAAAA ATGAACATCG ATACCGATAC CCAATGGGCA
4201 ACCTGGGAAG GCGTTCTGAA CTAATAACAA GCGAACGAAG CTTATCTGCA
4251 GGGTCAGCTG GGTAAACCGA AAGGCGAAGA TCAGCCGAAC AAGAAATACT
4301 ACGATCCCGC CGTATGGCTG CGTGCCGGTC AGACTTCGAT GATCGCTCGT
4351 CTGGAGAAAAG CATTCCAGGA ACTGAACGCG ATCGACGTTT TGTAAAGAGCT
4401 CCGTACCGGA TCCAATTccc GATCGTTCAA ACATTTGGCA ATAAAGTTTC
4451 TTAAGATTGA ATCCTGTTGC CGGTCTTGCG ATGATTATCA TATAATTTCT
4501 GTTGAATTAC GTTAAGCATG TAATAATTAA CATGTAATGC ATGACGTTAT

4551	TTATGAGATG	GGTTTTTATG	ATTAGAGTCC	CGCAATTATA	CATTTAATAC
4601	GCGATAGAAA	ACAAAATATA	GCGCGCAAAC	TAGGATAAAT	TATCGCGCGC
4651	GGTGTCACT	ATGTTACTAG	ATCGGGGATC	GATCCCCGGG	CGGCCGCCAC
4701	TCGAGTGGTG	GCCGCATCGA	TCGTGAAGTT	TCTCATCTAA	GCCCCCATTT
4751	GGACGTGAAT	GTAGACACGT	CGAAATAAAG	ATTTCCGAAT	TAGAATAATT
4801	TGTTTTATTGC	TTTCGCCTAT	AAATACGACG	GATCGTAATT	TGTCGTTTTA
4851	TCAAAATGTA	CTTTCATTTT	ATAATAACGC	TGCGGACATC	TACATTTTTG
4901	AATTGAAAAA	AAATTGGTAA	TTACTCTTTC	TTTTTCTCCA	TATTGACCAT
4951	CATACTCATT	GCTGATCCAT	GTAGATTTCC	CGGACATGAA	GCCATTTACA
5001	ATTGAATATA	TCCTGCCGCC	GCTGCCGCTT	TGCACCCGGT	GGAGCTTGCA
5051	TGTTGGTTTT	TACGCAGAAC	TGAGCCGGTT	AGGCAGATAA	TTTCCATTGA
5101	GAAGTGAGCC	ATGTGCACCT	TCCCCCAAC	ACGGTGAGCG	ACGGGGCAAC
5151	GGAGTGATCC	ACATGGGACT	TTTCTTAGCT	TGGCTGCCAT	TTTTGGGGTG
5201	AGGCCGTTCG	CGCGGGGCGC	CAGCTGGGGG	GATGGGAGGC	CCGCGTTACC
5251	GGGAGGGTTC	GAGAAGGGGG	GGCACCCCCC	TTCGGCGTGC	GCGGTCACGC
5301	GCCAGGGCGC	AGCCCTGGTT	AAAAACAAGG	TTTATAAATA	TTGGTTTAAA
5351	AGCAGGTAA	AAGACAGGTT	AGCGGTGGCC	GAAAAACGGG	CGGAAACCCT
5401	TGCAAATGCT	GGATTTTCTG	CCTGTGGACA	GCCCCTCAA	TGTCAATAGG
5451	TGCGCCCCCTC	ATCTGTCTATC	ACTCTGCCCC	TCAAGTGTCA	AGGATCGCGC
5501	CCCTCATCTG	TCAGTAGTCG	CGCCCCTCAA	GTGTCAATAC	CGCAGGGCAC
5551	TTATCCCCCAG	GCTTGTTCCAC	ATCATCTGTG	GGAAACTCGC	GTAATAATCAG
5601	GCGTTTTTCGC	CGATTTGCGA	GGCTGGCCAG	CTCCACGTCTG	CCGGCCGAAA
5651	TCGAGCCTGC	CCCTCATCTG	TCAACGCCGC	GCCGGGTGAG	TCGGCCCCCTC
5701	AAGTGTC AAC	GTCCGCCCTT	CATCTGTCTG	TGAGGGCCAA	GTTTTCCGCG
5751	TGGTATCCAC	AACGCCGGCG	GCCGGCCGCG	GTGTCTCGCA	CACGGCTTCG
5801	ACGGCGTTTT	TGGCGCGTTT	GCAGGGCCAT	AGACGGCCGC	CAGCCCAGCG
5851	GCGAGGGCAA	CCAGCCCGGT	GAGCGTCGGA	AAGGGTCGAT	CGACCGATGC
5901	CCTTGAGAGC	CTTCAACCCA	GTCAGCTCCT	TCCGGTGGGC	GCGGGGCATG
5951	ACTATCGTCG	CCGCACTTAT	GACTGTCTTC	TTTATCATGC	AACTCGTAGG
6001	ACAGGTGCCG	GCAGCGCTCT	GGGTCATTTT	CGGCGAGGAC	CGCTTTCGCT
6051	GGAGCGCGAC	GATGATCGGC	CTGTCTGCTT	CGGTATTCGG	AATCTTGCAC
6101	GCCCTCGCTC	AAGCCTTCGT	CACTGGTCCC	GCCACCAAAC	GTTTCGGCGA
6151	GAAGCAGGCC	ATTATCGCCG	GCATGGCGGC	CGACGCGCTG	GGCTACGTCT
6201	TGCTGGCGTT	CGCGACGCGA	GGCTGGATGG	CCTTCCCCAT	TATGATTCTT
6251	CTCGCTTCG	GCGGCATCGG	GATGCCCCGCG	TTGCAGGCCA	TGCTGTCCAG
6301	GCAGGTAGAT	GACGACCATC	AGGGACAGCT	TCAAGGATCG	CTCGCGGCTC
6351	TTACCAGCCT	AACTTCGATC	ACTGGACCGC	TGATCGTCAC	GGCGATTTAT
6401	GCCGCCCTCG	CGAGCACATG	GAACGGGTTG	GCATGGATTG	TAGGCGCCGC
6451	CCTATACCTT	GTCTGCCCTC	CCGCGTTGCG	TCGCGGTGCA	TGGAGCCGGG
6501	CCACCTCGAC	CTGAATGGAA	GCCGGCGGCA	CCTCGCTAAC	GGATTACCA
6551	CTCCAAGAAT	TGGAGCCAAT	CAATTCTTGC	GGAGAACTGT	GAATGCGCAA
6601	ACCAACCCTT	GGCAGAACAT	ATCCATCGCG	TCCGCCATCT	CCAGCAGCCG
6651	CACGCGGCGC	ATCTCGGGCA	GCGTTGGGTC	CTGGCCACGG	GTGCGCATGA
6701	TCGTGCTCCT	GTCGTTGAGG	ACCCGGCTAG	GCTGGCGGGG	TTGCCTTACT
6751	GGTTAGCAGA	ATGAATCACC	GATACGCGAG	CGAACGTGAA	GCGACTGCTG
6801	CTGCAAAACG	TCTGCGACCT	GAGCAACAAC	ATGAATGGTC	TTGCGTTTCC
6851	GTGTTTCGTA	AAGTCTGGAA	ACGCGGAAGT	CAGCGCCCTG	CACCATTATG
6901	TTCCGGATCT	GCATCGCAGG	ATGCTGCTGG	CTACCCTGTG	GAACACCTAC
6951	ATCTGTATTA	ACGAAGCGCT	GGCATTGACC	CTGAGTGATT	TTTCTCTGGT
7001	CCCGCCGCAT	CCATACCGCC	AGTTGTTTAC	CCTCACAACG	TTCCAGTAAC
7051	CGGGCATGTT	CATCATCAGT	AACCCGTATC	GTGAGCATCC	TCTCTCGTTT
7101	CATCGGTATC	ATTACCCCCA	TGAACAGAAA	TTCCCCCTTA	CACGGAGGCA
7151	TCAAGTGACC	AAACAGGAAA	AAACCGCCCT	TAACATGGCC	CGCTTTATCA
7201	GAAGCCAGAC	ATTAACGCTT	CTGGAGAAAC	TCAACGAGCT	GGACGCGGAT

7251 GAACAGGCAG ACATCTGTGA ATCGCTTCAC GACCACGCTG ATGAGCTTTA
7301 CCGCAGCTGC CTCGCGCGTT TCGGTGATGA CGGTGAAAAC CTCTGACACA
7351 TGCAGCTCCC GGAGACGGTC ACAGCTTGTC TGTAAGCGGA TGCCGGGAGC
7401 AGACAAGCCC GTCAGGGCGC GTCAGCGGGT GTTGGCGGGT GTCGGGGCGC
7451 AGCCATGACC CAGTCACGTA GCGATAGCGG AGTGTATACT GGCTTAACTA
7501 TGCGGCATCA GAGCAGATTG TACTGAGAGT GCACCATATG CGGTGTGAAA
7551 TACCGCACAG ATGCGTAAGG AGAAAATACC GCATCAGGCG CTCTTCCGCT
7601 TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCAGCGGGT
7651 ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA TCAGGGGATA
7701 ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT
7751 AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC CCCCTGACGA
7801 GCATCAGAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC
7851 TATAAAGATA CCAGGCGTTT CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT
7901 GTTCCGACCC TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG
7951 AAGCGTGGCG CTTTCTCATA GCTCACGCTG TAGGTATCTC AGTTCGGTGT
8001 AGGTCGTTCC CTCCAAGCTG GGCTGTGTGC ACGAACCCCC CGTTCAGCCC
8051 GACCGCTGCG CTTTATCCGG TAACTATCGT CTTGAGTCCA ACCCGGTAAG
8101 ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG
8151 CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC
8201 GGCTACACTA GAAGGACAGT ATTTGGTATC TGCGCTCTGC TGAAGCCAGT
8251 TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG
8301 CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC AGCAGATTAC GCGCAGAAAA
8351 AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT CTGACGCTCA
8401 GTGGAACGAA AACTCACGTT AAGGGATTTT GGTCATGAGA TTATCAAAAA
8451 GGATCTTCAC CTAGATCCTT TTAAAT'AAA AATGAAGTTT TAAATCAATC
8501 TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG
8551 TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT
8601 GACTCCCCGT CGTGTAGATA ACTACGATAC GGGAGGGCTT ACCATCTGGC
8651 CCCAGTGCTG CAATGATACC GCGAGACCCA CGCTCACCAG CTCCAGATTT
8701 ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG
8751 CAACTTTATC CGCCTCCATC CAGTCTATTA ATTGTTGCCG GGAAGCTAGA
8801 GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG CCATTGCTGC
8851 AGGTCGGGAG CACAGGATGA CGCCTAACAA TTCATTCAAG CCGACACCGC
8901 TTCGCGGCGC GGCTTAATTC AGGAGTTAAA CATCATGAGG GAAGCGGTGA
8951 TCGCCGAAGT ATCGACTCAA CTATCAGAGG TAGTTGGCGT CATCGAGCGC
9001 CATCTCGAAC CGACGTTGCT GGCCGTACAT TTGTACGGCT CCGCAGTGGA
9051 TGGCGGCCCTG AAGCCACACA GTGATATTGA TTTGCTGGTT ACGGTGACCG
9101 TAAGGCTTGA TGAAACAACG CGGCGAGCTT TGATCAACGA CCTTTTGGA
9151 ACTTCGGCTT CCCCTGGAGA GAGCGAGATT CTCCGCGCTG TAGAAGTCAC
9201 CATTGTTGTG CACGACGACA TCATTCCGTG GCGTTATCCA GCTAAGCGCG
9251 AACTGCAATT TGGAGAATGG CAGCGCAATG ACATTCTTGC AGGTATCTTC
9301 GAGCCAGCCA CGATCGACAT TGATCTGGCT ATCTTGCTGA CAAAAGCAAG
9351 AGAACATAGC GTTGCCCTTG TAGGTCCAGC GGCGGAGGAA CTCTTTGATC
9401 CGGTTCCCTGA ACAGGATCTA TTTGAGGCGC TAAATGAAAC CTTAACGCTA
9451 TGGAAC'TCGC CGCCCGACTG GGCTGGCGAT GAGCGAAATG TAGTGCTTAC
9501 GTTGTCCCGC ATTTGGTACA GCGCAGTAAC CGGCAAAATC GCGCCGAAGG
9551 ATGTCGCTGC CGACTGGGCA ATGGAGCGCC TGCCGGCCCA GTATCAGCCC
9601 GTCATACTTG AAGCTAGGCA GGCTTATCTT GGACAAGAAG ATCGCTTGGC
9651 CTCGCGCGCA GATCAGTTGG AAGAATTTGT TCACTACGTG AAAGCGGAGA
9701 TCACCAAGGT AGTCGGCAAA TAATGTCTAA CAATTCGTTT AAGCCGACGC
9751 CGCTTCGCGG CGCGGCTTAA CTCAAGCGTT AGATGCTGCA GGCATCGTGG
9801 TGTCACGCTC GTCGTTTGGT ATGGCTTCAT TCAGCTCCGG TTCCCAACGA
9851 TCAAGGCGAG TTACATGATC CCCCATGTTG TGCAAAAAAG CGGTTAGCTC
9901 C'TTCGGTCCT CCGATCGAGG ATTTTTCGGC GCTGCGCTAC GTCCGCKACC

9951 GCGTTGAGGG ATCAAGCCAC AGCAGCCCAC TCGACCTCTA GCCGACCCAG
 10001 ACGAGCCAAG GGATCTTTTT GGAATGCTGC TCCGTCGTCA GGCTTTCCGA
 10051 CGTTTGGGTG GTTGAACAGA AGTCATTATC GTACGGAATG CCAAGCACTC
 10101 CCGAGGGGAA CCCTGTGGTT GGCATGCACA TACAAATGGA CGAACGGATA
 10151 AACCTTTTCA CGCCCTTTTA AATATCCGTT ATTCTAATAA ACGCTCTTTT
 10201 CTCTTAGGTT TACCCGCCAA TATATCCTGT CAAACACTGA TAGTTTAAAC
 10251 TGAAGGCGGG AAACGACAAT CTGATCCCCA TCAAGCTTGA GCTCAGGATT
 10301 TAGCAGCATT CCAGATTGGG TTCAATCAAC AAGGTACGAG CCATATCACT
 10351 TTATTCAAAT TGGTATCGCC AAAACCAAGA AGGAACTCCC ATCCTCAAAG
 10401 GTTTGTAAGG AAGAATTCTC AGTCCAAAGC CTCAACAAGG TCAGGGTACA
 10451 GAGTCTCCAA ACCATTAGCC AAAAGCTACA GGAGATCAAT GAAGAATCTT
 10501 CAATCAAAGT AAACACTGT TCCAGCACAT GCATCATGGT CAGTAAGTTT
 10551 CAGAAAAAGA CATCCACCGA AGACTTAAAG TTAGTGGGCA TCTTTGAAAG
 10601 TAATCTTGTC AACATCGAGC AGCTGGCTTG TGGGGACCAG ACAAAAAAGG
 10651 AATGGTGCAG AATTGTTAGG CGCACCTACC AAAAGCATCT TTGCCTTTAT
 10701 TGCAAAGATA AAGCAGATTC CTCTAGTACA AGTGGGGAAC AAAATAACGT
 10751 GGAAAAGAGC TGTCCTGACA GCCCACTCAC TAATGCGTAT GACGAACGCA
 10801 GTGACGACCA CAAAAGAATT CCCTCTATAT AAGAAGGCAT TCATTCCCAT
 10851 TTGAAGGATC ATCAGATACT GAACCAATCC TTCTAGAAGA TCTAAGCTTA
 10901 T